A Publication of the American Psychological Association

The Journal of Abnormal and Social Psychology

Founded by Morton Prince, 1906

CONTENTS

Articles

The Nature of Hypnosis: Artifact and Essence 277
Martin T. Orne

Drive Strength and Symptom Manifestations 300
J. R. Wittenborn and Nathan S. Kline

Verbal Rate, Content, and Adjustment Ratings in Experimentally Structured Interviews 305
Frederick H. Kanfer

Hostility in Rorschach Content and Overt Aggressive Behavior 312
Alan P. Towbin

The Generality of Social Perception Scores 317
Mohammed Y. Qureshi

Subjective Certainty and Resistance to Change 323
Donald S. Boomer

Continued on inside front cover

This is the Last Issue of Volume 58.
Title Page and Contents for the Volume
Appear Herein.

Volume 58 MAY 1959 Number 3

CONTENTS—*Continued*

The Effects on Opinions of a Change in Scale Judgments 329
Walter Weiss

Structural Analysis of Dreams: Clues to Perceptual Style 335
Eernice T. Eidusons

Autonomic Responses and Their Relationship to Race Attitudes 340
Frank R. Westie and Melvin L. De Fleur

Effect of Changes in Self-Esteem Upon Impulsiveness and Deliberation in Making Judgments 348
James E. Dittes

Stability and Validity of Achievement Fantasy 357
Jerome Kagan and Howard A. Moss

Need Affiliation and Frequency of Four Types of Communication 365
John B. Lansing and Roger W. Heyns

Re-evaluation of Choice Alternatives as a Function of Their Number and Qualitative Similarity 373
Jack W. Brehm and Arthur R. Cohen

Increasing Cognitive Dissonance by a *Fait Accompli* 379
Jack W. Brehm

Choice and Chance Relative Deprivation as Determinants of Cognitive Dissonance 383
Jack W. Brehm and Arthur R. Cohen

Attitudinal Effects of Choice in Exposure to Counterpropaganda 388
Arthur R. Cohen, Herbert I. Terry, and Charles B. Jones

Evaluation of Therapy by Self-Report: A Paradox 392
Jane Loevinger and Abel Osserio

CRITIQUE AND NOTES

Displacement of Preference as a Function of Avoidance-Avoidance Conflict .. 395
Hava Bonné Gewirtz

Personality, Perceptual Defense, and Stereoscopic Perception 398
John M. Davis

The Relationship Between GSR and Speech Disturbances in Psychotherapy .. 402
David M. Panek and Barclay Martin

Some Factors in the Selection of Leaders by Members of Small Groups 406
John P. Kirscht, Thomas M. Lodahl, and Mason Haire

Yielding, Authoritarianism, and Authoritarian Ideology Regarding Groups .. 408
Eugene B. Nadler

Volume 58 MAY, 1959 Number 3

The Journal of Abnormal and Social Psychology

Editor: M. BREWSTER SMITH, *New York University*

Associate Editor: E. J. SHOBEN, JR., *Teachers College, Columbia University*

CONSULTING EDITORS

GORDON W. ALLPORT
Harvard University

URIE BRONFENBRENNER
Cornell University

ROGER BROWN
Massachusetts Institute of Technology

DONALD T. CAMPBELL
Northwestern University

LAUNOR F. CARTER
System Development Corporation

RICHARD CHRISTIE
Columbia University

MORTON DEUTSCH
Bell Telephone Laboratories

I. E. FARBER
State University of Iowa

DONALD W. FISKE
University of Chicago

ROGER W. HEYNS
University of Michigan

ROBERT R. HOLT
New York University

IRVING L. JANIS
Yale University

HAROLD E. JONES
University of California

HAROLD H. KELLEY
University of Minnesota

LEONARD S. KOGAN
Community Service Society of New York

DANIEL J. LEVINSON
Harvard Medical School

DAVID C. MCCLELLAND
Harvard University

LEO POSTMAN
University of California

ELIOT H. RODNICK
Duke University

NEVITT SANFORD
University of California

SEYMOUR SARASON
Yale University

RICHARD L. SOLOMON
Harvard University

RENATO TAGIURI
Harvard University

PUBLISHED BY

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

The JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY is devoted to both abnormal and social psychology. Its emphasis is on basic research and theory rather than the techniques and arts of practice. Abnormal psychology is broadly defined to include papers contributing to fundamental knowledge of the pathology, dynamics, and development of personality or individual behavior, including deterioration with age and disease. Articles concerned with psychodiagnostic techniques are evaluated with respect to their contribution to an understanding of the psychological principles of diagnosis, those concerned with psychotherapy are judged in terms of their contribution to an understanding of the therapeutic process. Case reports are considered in terms of their heuristic value. Reports which serve to pose or to clarify important theoretical problems, or which promise to be useful in teaching, are often published. From the social area this JOURNAL gives preference to papers contributing to basic knowledge of interpersonal relations, and of group influences on the pathology, dynamics, and development of individual behavior.

The JOURNAL is issued bimonthly in January, March, May, July, September, and November, two volumes per year. The two volumes average 832 pages. The subscription price per year is $16.00; foreign, $16.50. Single issues are $3.00.

Original manuscripts should be sent to the Editor: M. BREWSTER SMITH, Graduate Department of Psychology, New York University, Washington Square, New York 3, N. Y.

NOTE CHANGE OF EDITORIAL ADDRESS. After June 30, manuscripts and other communications should be addressed to the Editor, M. Brewster Smith, Department of Psychology, University of California, Berkeley 4, California.

ARTHUR C. HOFFMAN
Managing Editor

HELEN ORR
Promotion Manager

BARBARA CUMMINGS
Editorial Assistant

Subscriptions, address changes, and business communications should be sent to:

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1333 Sixteenth Street, N.W.
Washington 6, D. C.

Changes of address must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
Mt. Royal and Guilford Aves., Baltimore 2, Md.
and 1333 Sixteenth Street N.W., Washington 6, D.C.

Second-class postage paid at Baltimore, Maryland, and at additional mailing offices

Copyright 1959 by the American Psychological Association, Inc.

THE NATURE OF HYPNOSIS: ARTIFACT AND ESSENCE[1]

MARTIN T. ORNE

Harvard University and Massachusetts Mental Health Center

THE most meaningful present-day theories of hypnosis interpret hypnotic phenomena along three major lines: (*a*) desire on the part of the subject to play the role of a hypnotized subject (Sarbin, 1950; White, 1941), (*b*) increase in suggestibility (Hull, 1933), and (*c*) a further less well-defined category that is called by White "an altered state of consciousness" and by others, "cortical inhibition" (Pavlov, 1923), dissociation (Weitzenhoffer, 1953), etc. depending on their theoretical orientations.

The heuristic model of hypnosis that underlies this paper incorporates these three aspects. One of the hypotheses of the paper holds that much hypnotic behavior results from the subject's conception of the role of the hypnotic subject as determined by past experience and learning, and by explicit and implicit cues provided by the hypnotist and the situation. These varied role conceptions appear to be the source of most if not all of the inconstant patterns of behavior seen in the hypnotic state.

An increase in suggestibility may be viewed as an increase in motivation to conform to the wishes of the hypnotist. A second basic hypothesis to be tested thus proposes that, although increased motivation may be a constant accompaniment of the trance state, such increased motivation is by no means a phenomenon unique to hypnosis but can be seen to operate in other experimental and life situations with equal force.

By experimentally controlling these two elements, role-playing and increased motivation, it is possible to investigate their sufficiency for explaining all aspects of the trance state and the extent to which still other concepts, such as an altered state of consciousness, are required.

The third aspect of hypnosis, the altered state of consciousness, presents the greatest problem for investigation, yet it has been felt necessary to include the concept in all attempts to explain the phenomenon. This residual aspect, which remains after increased motivation and role-playing are accounted for, may be regarded as the "essence" of hypnosis, with reference to which increased motivation and role-playing appear as artifacts.

Three related experiments are presented. The first is devoted to the effects of "role-play artifact" on the manifestations of hypnosis commonly seen clinically. It demonstrates that much of the complex phenomenon which we call hypnosis may result from (*a*) the subject's preconceptions of what hypnosis is, (*b*) implicit cues by the hypnotist as to what he thinks it should be, and (*c*) the particular techniques of trance induction. The second experiment demonstrates an aspect of role-play artifact that is introduced by a concrete experimental situation. It investigates cues that an experimental design may give about the role the subject is expected to play and demonstrates that in some instances an experimental result may more reasonably be accounted for on this basis than by invoking "trance effects." The third experiment is concerned with the effect of "motivation artifact" upon performance. It examines the claims of increased physical capacity in hypnosis and tests the hypothesis that this may be accounted for by increased motivation.

Table 1 gives a schematic representation of the author's working model of the hypnotic state.

[1] This investigation was supported in part by a postdoctoral research fellowship from the Public Health Service.

I would like to thank Robert W. White for his encouragement and guidance and Milton Greenblatt for his support and his cooperation in the use of the facilities of the Massachusetts Mental Health Center. I am greatly indebted to Donald O'Connell, for his help in the pilot studies, and Ronald Shor and Theodore X. Barber for their invaluable assistance in running the first two experiments reported here. In addition, thanks are due Peter D. Watson for his aid in preparing and editing the manuscript. Finally, I would like to express my appreciation to Abraham H. Maslow of Brandeis University, Philip Nogee of Boston University, William McGill of the Massachusetts Institute of Technology, E. G. Boring and Ray Hyman of Harvard University, and the respective departments of psychology for making their classes available to me.

TABLE 1

SCHEMATIC REPRESENTATION OF A WORKING MODEL OF HYPNOSIS

Situation of trance induction	"Role-Play Artifact" (cognitive component)	+	"Increased Motivation Artifact" (conative component)	+	Essence of Trance
Creation of situation to maximize: 1. Desirability of entering trance 2. Expectation that trance can be achieved 3. Respect and trust for operator 4. Restriction of extraneous stimuli 5. Focusing of attention	1. Expectations of *S*s a. preconceptions b. cues from trance induction 2. Cues from Experimenter a. explicit b. implicit 3. Cues from experimental situation		The sources of increased motivation are not defined They represent a major area of future inquiry Probably some aspects will prove to be a component of "essence"		Uncertain
All techniques have the further qualities of: 1. Concrete suggestions in vivid simple language 2. "Suggestions" utilizing the perception of subjective events as their basis 3. Suggestions of gradually increasing difficulty to insure successful responses 4. Praising (rewarding) explicitly or implicitly the subject's positive responses					

PRECONCEPTIONS OF HYPNOSIS AND THEIR EFFECT ON TRANCE MANIFESTATIONS

The states induced by Mesmer (Binet & Féré, 1888; Boring, 1950), Coué (1922, p. 83), Wells (1923), Schilder (1956), and others are all hypnosis, yet their descriptions of how hypnosis characteristically manifests itself are very different. The common characteristics of these varied states that bring them all under the heading of "hypnosis" would appear to include: posthypnotic amnesia, apparent inability to use a given motor system when a functional paralysis is suggested, various sensory illusions including positive and negative hallucinations of all sensory modalities, apparent memory disturbances or improvements as well as reported increased control over autonomic nervous system functions. Whether all of these phenomena are necessarily part of hypnotic behavior will be discussed below. In any event, hypnosis is evidently characterized by the ability of the subject (*S*) in this special state to experience changes that are not normally found in response to similar cues in everyday life.

What, then, determines the particular trance manifestations that an *S* shows on entering hypnosis? In terms of the model presented here, the answer may lie in role-play artifact. From this viewpoint, *S*s who enter trance are motivated to play the role of the hypnotized *S*, and the precise manifestations of this role depend upon their perception of what it entails. Behavior of the *S* in trance is then determined by the *S*'s preconceptions about how a hypnotic *S* acts, and the cues, both explicit and implicit, as to the desired behavior which the hypnotist communicates in the process of trance induction.

To test this hypothesis that conceptions about hypnosis held prior to entering the hypnotic state affect an *S*'s trance behavior, a pilot study and a main experiment were conducted in which volunteer *S*s were given the erroneous prior impression that catalepsy of the dominant hand (with the other hand flaccid) is a typical feature of hypnosis. This behavioral item was chosen because it satisfied a number of criteria. It is sufficiently unusual to have been reported had it ever been observed as a spontaneous characteristic of hypnosis; it is easily recognizable so that judgments of its presence or absence are unequivocal; and it is sufficiently plausible as a charac-

teristic associated with hypnosis that it would be accepted as such by the *S* population.

Especial care was exercised to eliminate possible effects of the bias of the experimenter by making it impossible for him to influence the results. It is easy to suggest to an *S* by implicit cues that he manifest catalepsy as part of the hypnotic state. Perhaps catalepsy of one hand might also be suggested during induction of the trance. Selection of catalepsy of the *dominant* hand avoids this possibility, as the experimenter had no way of knowing whether the subject was right- or left-handed until he asked for this information after the data on catalepsy had been gathered.

Pilot Study

An introductory psychology class at the Massachusetts Institute of Technology was given a lecture on hypnosis. Prior to the lecture, and without the knowledge of the class, two students had been hypnotized and given the posthypnotic suggestion that upon entering the trance subsequently, they would manifest catalepsy of one hand, the dominant hand. One student was right-handed and one student was left-handed. The class was then given a 25-minute lecture on the nature of the hypnotic state, at which point volunteers were called for in order to demonstrate the phenomenon. Of the 11 students who volunteered, the two who had been previously hypnotized were selected in a fashion that appeared random. They were again placed in trance, in a manner that appeared to be the initial trance induction, and simple trance phenomena were demonstrated, including one-handed catalepsy. Attention was called to the fact that the right-handed student had catalepsy of the right hand, and the left-handed student had catalepsy of the left hand. Immediately following this procedure, three more students from the same group of volunteers, who had not been hypnotized previously, were placed in trance.

A class of psychology students at Harvard were subsequently given the same kind of a lecture and demonstration, following which four *S*s were hypnotized and tested for one-handed catalepsy.

All three of the M.I.T. experimental *S*s gave good trance results, and all showed catalepsy of the dominant hand. One *S* was left-handed. Of the four Harvard students who were hypnotized immediately after observing three demonstration *S*s with catalepsy of the dominant hand, three manifested catalepsy of the dominant hand and one, catalepsy of both hands. All *S*s were right-handed.

Main Study

In order to make it impossible for the experimenter to communicate his desire that the *S* demonstrate unilateral catalepsy, the main study was performed in a rigorous "blind" fashion. In this instance matched classes were used, each of which had received a lecture and demonstration of hypnosis. In one class the hypnotic demonstration included catalepsy of the dominant hand, while in the other this was omitted. The *S*s were then tested in small groups, with members of both groups mixed randomly. The experimenter thus had no way of knowing which subjects should manifest one-handed catalepsy.

Procedure

The procedure of the pilot experiment was repeated with members of the introductory psychology course at Boston University with the inclusion of the control group. Instead of asking for volunteers, three *S*s were employed who were introduced to the class as having taken part in prior research. The same three *S*s were used for both sections of the course, to which essentially identical lectures were given. The demonstrations differed only in that in one section the three *S*s manifested unilateral catalepsy, while in the other section this was not demonstrated. No students from either class were hypnotized at that time. Volunteers were solicited and subsequently tested in such a way that the experimenter had no way of telling which lecture they had attended until after the completion of the experiment.[2] All but two *S*s were tested by an experimenter who was not at the lectures.

Trance depth was rated by the experimenter and an observer. The degree of consensus was high and in no case was there more than a one point difference. In case of disagreement both ratings are recorded. The ratings are rough clinical estimates based on the phenomena

[2] One of these *S*s was tested the evening of the lecture. The remaining *S*s were tested approximately one month after the lecture.

TABLE 2
TRANCE BEHAVIOR IN THE EXPERIMENTAL AND CONTROL GROUPS

Subject	Catalepsy		Trance Depth	Dominant Hand
	R.	L.		
Experimental group				
1. M.S.	+	0	4	R.
2. M.K.	+	+	4	R.
3. R.L.	+	0	4	R.
4. C.L.	0	0	2	R.
5. S.T.	0	0	3	R.
6. A.L.	0	+	3	L.
7. O.B.	+	0	3	R.
8. S.R.	+	+	5	R.
9. B.T.	+	0	4	R.[a,b]
Control group				
1. D.L.	0	0	4	R.
2. W.O.	0	0	4	R.
3. M.R.	+	+	3	R.
4. L.P.	+	+	3	L.
5. B.Z.	0	0	3	R.
6. L.V.	0	0	4–5	R.
7. M.O.	+	+	3	R.
8. A.T.	0	0	3	L.[b]
9. W.M.	0	0	1–2	R.

[a] This *S* was tested the evening of the lecture when he appeared unannounced along with a group of *S*s who had previously volunteered. The experimenter did not know which class the *S* had attended until after the experiment was over.

[b] *S*s tested by author. I was not aware of which class these *S*s had attended, in fact, I did not know until subsequently that they had been at the lectures.

which could be elicited from the *S*s. A rating of 1 indicated no response; 2 implied eye-closure and only partial hand levitation without a positive response to "challenge" suggestions, i.e., you cannot open your eyes, or you cannot bend your elbow; 3 referred to positive responses to all challenge suggestions but inability to achieve hallucinations or posthypnotic phenomena; 4 was used to denote those *S*s who responded to suggested hallucinations, gave simple posthypnotic phenomena, but did not achieve a good post-hypnotic amnesia; 5 referred to "somnambulists" who could achieve all hypnotic phenomena easily, including complete amnesia.

Results

Of the nine *S*s in the experimental group, five showed catalepsy of the dominant hand. Two showed catalepsy of both hands, and two showed no catalepsy. None of the control group showed catalepsy of the dominant hand, but three out of the nine *S*s showed catalepsy of both hands. Table 2 gives a summary of the findings.

Discussion

The item of behavior that was used is known not to occur spontaneously; its occurrence is significant if it is found at all. The results of the pilot and main experiments may be regarded as confirming the hypothesis that trance behavior is affected by the individual's conceptions about hypnosis held prior to entering the hypnotic state.

It would not be expected that all *S*s would show this behavior. No truly naive *S* population is available, and many of the *S*s had observed hypnosis prior to the demonstration. Some *S*s should therefore have sufficient prior information to have formed very strong conceptions unlikely to be altered by the relatively mild attempt to manipulate these ideas experimentally.

That three of the nine *S*s in the control group spontaneously manifested catalepsy of both hands is readily understood in view of the repeated testing for catalepsy, which they apparently interpreted as a cue to manifest the behavior. None of the control *S*s, it should be emphasized, manifested unilateral catalepsy, indicating that no such desire was communicated by the hypnotist to the *S*.

This study has demonstrated for a single behavior item that trance behavior is affected by individual preconceptions about hypnosis. The results can be extrapolated to account for the apparently fixed qualities, not stemming from cues given by the hypnotist, that are reported in practically all present-day descriptions of hypnosis.

Thanks to the media of mass communication, it is relatively easy for a particular view of hypnosis to have gained wide currency and thus be found as a part of the general knowledge in which the *S*s share. Such novels as *Mario and the Magician* (Mann, 1931) and *Trilby* (DuMaurier, 1895) have had very wide audiences and are known indirectly to almost all members of our culture. Uncounted articles and features about hypnosis have been disseminated to all levels of society. The picture of hypnosis that emerges in all of these is that of a passive *S* in a sleeplike state who has amnesia for the events occurring in hypnosis, and responds only to the hypnotist's sugges-

tions. According to Dorcus, Brintmall, and Case (1941), 79% of the student sample that they studied accepted hypnosis as possible, 71% had discussed hypnosis with someone, 54% had read about it, and 29% had actually seen a hypnotic trance at one time during their lives.

In the context of group tests for "suggestibility," in order to screen *S*s, the investigators asked 57 students in elementary psychology courses: "Have you observed any other demonstrations of hypnosis; if so, where and when?" and "What have you read about hypnosis?" Only 12 *S*s denied both having read about hypnosis and having had any chance to see the phenomenon previously; 18 *S*s had seen hypnosis demonstrated in some form, and 23 had somehow read about it.

In the context of the questionnaires used in the above studies, "having read about hypnosis" meant specific reading in the scientific sense. In questioning well over 200 student *S*s about their knowledge of hypnosis, the author failed to find one who did not have a very clear-cut notion about the nature of hypnosis, and who could not define the trance in a fashion similar to that found in dictionaries. Furthermore, they had all read something about hypnosis and could recall having done so, once it was made clear that this included nonscientific sources. The normal *S* population thus knows the meaning of the word hypnosis prior to taking part in any study.

CUES IMPLICIT IN AN EXPERIMENTAL DESIGN

An *S* participating in an experiment is aware that his responses are being recorded for specific purposes—that there is a *raison d'être* for the experiment—and he frequently has some idea of what these purposes are. How this knowledge affects the *S*'s behavior depends upon the motivational structure that he brings to the experimental situation. The participation of the college student volunteer in psychological studies is usually due, not to the relatively low monetary remuneration but, rather, to his interest in taking part in scientific research, which in turn is likely to be based, at least in part, on a desire to further "progress in science" by his participation. Since the experimenter is perceived as knowing what he is doing, furthering "progress in science" may well be equated with "making the experiment work" or, in more sophisticated terms, having his individual performance support the hypothesis of the experiment. Thus, when the *S* is motivated to comply with the wishes of the experimenter, his responses are readily influenced by what he perceives to be the basic hypothesis of the experiment.

Typically, the experimenter's hypotheses are not stated explicitly to the *S* because of the very considerations just mentioned. But unstated hypotheses may be conveyed implicitly by the experimental procedure itself, through what will be called here the "demand characteristics of the experimental situation." It should be understood that a person may fail to perceive fairly clear demand characteristics either because of lack of past experience or because of an inability to generalize from it.

Demand characteristics thus conceived appear central to much psychological work. Experimental situations vary widely in the extent to which they convey the purpose and the hypothesis of the experimenter. If an *S* can describe a hypothesis being tested, of which he is supposedly unaware, the experimental arrangements have significant demand characteristics. The obvious way to test for their presence is to ask the *S* about his perception of the experiment and its purpose. Usually, however, *S*s are reticent about revealing their notions about the purpose of the experiment.

It is reasonable to assume that the student *S* population has some sophistication in regard to the philosophy of experimentation. They are aware that if an *S* is not told the purpose of an experiment he ought to remain naive in regard to it, lest his knowledge influence his performance. At the same time they understand the necessity for an experimental *S* to be "honest" in his response to the experimental situation and to questions about it. For these reasons, *S*s are motivated to avoid recognizing explicitly the purpose of an experiment even though it may be clearly communicated by its design. Thus, the response to the direct question "What do you think this is about?" tends to be "I don't know." The *S*'s behavior may nevertheless clearly betray an implicit aware-

ness of the relevant factors, and he may even verbalize them after the experiment in a "bull session" with his friends. We deal, therefore, with "knowledge" not readily available to consciousness which must be elicited in a clinical fashion. As in the case of other such material, the boundaries of consciousness may be expected to vary with the situation. When, however, a clinical approach is used in an inquiry and the *S* is pressed, one may be amazed—or horrified—by the *S*'s ability to formulate one's hypotheses in a lucid and at times highly sophisticated fashion. Unfortunately, the so-called inquiry is usually a most casual procedure.

While the demand characteristics of experimental situations probably have wider significance than is generally recognized, they are particularly significant for hypnotic experiments. Hypnotic *S*s tend to be particularly cooperative, almost eager participants. Furthermore, one of the assumptions of the present research for which there is extensive observational support is that the hypnotic state as such increases the motivation of the *S* to comply with the wishes ("suggestions")—both explicit and implicit—of the experimenter. The extent to which compliance can take place depends upon the demand characteristics in the experimental situation. The usual problem of demand characteristics (difficult enough to control in other fields of psychology because of the unconscious cooperation between *S* and experimenter) is thus compounded in hypnotic research.

In order to investigate the influence of the demand characteristics of an experimental procedure, a recent study (Ashley, Harper, & Runyon, 1951) was repeated with minor variations to be described. This experiment attempts to demonstrate a further dimension of the Bruner-Goodman (1947) effect, which has been the center of major controversy in recent years. Bruner and Goodman's basic tenet was that the perceiver's values alter his perception. There is no question that the perceiver's *previous experiences* may affect perception. A dispute, however, centers about whether *values* as such are significant variables affecting perception.

In order to show "clearly and unequivocally that the perceiver can contribute to the organization of his perception in a structured stimulus-situation," Ashley, Harper, and Runyon (1951) argue it would be necessary to have a special situation. They state: "The Bruner and Goodman type of experiment would do this *if* the rich group and the poor group were identical in every other respect—in terms of their experience with money, their life histories, their physiological conditions, in short, if the sole difference between the two groups was that only one group had the psychological organization . . . of rich people and the other group the psychological organization of poor people." They go on to say: "Actually for our problem, it is irrelevant whether the *S*s are economically as well as psychologically rich or poor, or whether they are only psychologically rich or poor. In either case, a difference in performance of the two groups would reflect a difference in the perception due to the psychological organization of the perceivers" (p. 565).

In order to obtain two groups identical in every respect but for their perception of their economic status, they used hypnosis. While the *S* was in trance, artificial life histories were induced—one rich and one poor—each followed by induced amnesia. In essence, then, they view the situation as if two identically matched groups were available—one rich, and one poor. It is assumed that because amnesia was induced for the preceding state, the *S* is again naive and that the only difference is in respect to his perceived economic status.

The final sentences of their rationale are particularly interesting. "Even though we do not know fully what happens when we hypnotize a person, if we do hypnotize him and tell him he is rich and he behaves in one way in the coin-matching situation, and then, a few moments later, we tell him he is now poor and he behaves in another way, *we can conclude that the observed difference is due to a change in his psychological organization*" (Ashley et al., 1951, p. 565).[3] The authors in fact conclude from their data that the psychological organization (including the wants, needs, interests, attitudes, and values) of the person contributes to the figural organizations of his perceptions.

It is unquestionably true that observed differences in coin-size judgments are due to

[3] Italics mine.

changes in psychological organization. The question with which we are concerned, however, is whether these changes in psychological organization relate to the actual experiencing of the feelings of being rich or poor, or whether they reflect the demand characteristics of the experimental procedure. The hypothesis to be tested is that the demand characteristics of the experiment are largely responsible for the results obtained by Ashley et al. (1951).

Disregarding the theoretical framework of the study, this is what actually took place: An individual was told—in hypnosis—that he was very poor, then—again with amnesia in hypnosis—that he was very rich and, subsequently, with another hypnotically induced amnesia, that he was himself. In each of these states he was required to make a series of coin-size judgments. The authors' interpretation rests largely on the assumption that hypnotic amnesia is truly the same as not knowing. Granted this, one would be justified in ignoring the fact that the procedure of coin-size estimation is repeated and that economic status is hypnotically induced. However, data are available that lead one to question this assumption.

One of the few specific experiments dealing with posthypnotic amnesia directly is a study by Strickler (1929), who compared the relearning of nonsense material in the posthypnotic state with induced amnesia with the learning time required for the material not previously learned. He concludes that "the posthypnotic amnesia ordinarily met with, which appears superficially to be a complete wiping-out of memory, is by no means complete."

Even more relevant are the data obtainable in hypnotic age-regression. Here we are dealing with an induced amnesia in hypnosis for what purports to be all material learned after a given age. All studies of hypnotic age-regression have shown that some material persists no matter how "real" the regression appears.

In the investigator's prior work (Orne, 1951), it was possible to show that an individual regressed to age six was able to comprehend English, though he himself pointed out in German that he could not understand it. Historically, the *S* was unable to understand English at age six. Another *S* could spell without error "I am conducting an experiment which will assess my psychological capacities." Another was able to give the square root of four, and so on. Furthermore, if we test for amnesia in a more subtle fashion, it is easy to demonstrate in the wake state or in trance that no true ablation of the material for which the *S* has amnesia exists, despite his subjective feeling of being unable to remember.

The fallacy of the assumption that knowledge for which the *S* has amnesia does not influence his behavior can be seen in any posthypnotic suggestion. The *S* firmly denies recall yet assiduously persists in the suggested behavior. The phenomenon is well known in response to an explicit cue; it would seem rather absurd to deny it in response to an implicit one.

A pilot study was therefore conducted that replicated all essential characteristics of the Ashley, Harper, and Runyon experiment, with the addition, however, of a careful inquiry after the completion of the experiment. The procedure was patterned after the inquiries commonly performed as part of the Rorschach test, which seek answers to a series of questions without providing the *S* with a cue as to the answers expected. 1. The subject's perception of the experimental task was elicited by a general question, "What do you think this experiment was about?" 2. The *S*'s perception of the purpose of the investigation was elicited by questions such as "What do you think this experiment is trying to prove or demonstrate?" 3. *S*'s perception of the experimenter's hypothesis was elicited by direct questioning, with such questions as "What do you think I hope to find?" 4. The *S* was also asked about his own hypothesis concerning the study—what he, on the basis of what he knew about the experiment, would predict the results to be. 5. The final question related to his beliefs about his own performance with the question, "What do you think your experimental behavior demonstrates?"

The following hypotheses were formulated:

1. The subject in an experiment is usually able to express some demand characteristics of the procedure, if careful inquiry is conducted and his initial resistance is penetrated in a clinical fashion.

2. The majority of subjects may perceive the same demand characteristics in the experi-

ment and these may be the same as the hypothesis being tested.

3. These demand characteristics rather than the experimental variables may be the major determinant of the subject's behavior.

a. If the majority of subjects perceive the same demand characteristics, then subjects who fail to perceive them should not show the behavior characteristic of the group.

b. If the demand characteristics are the determinant of subjects' behavior, it is possible for an experimental design that omits a crucial aspect of the original independent variable to elicit similar responses to the extent that the same demand characteristics are present.

Pilot Study

The pilot study was designed to test the first two hypotheses.

Procedure

The Ashley, Harper, and Runyon study was repeated in all essential details with four undergraduate *S*s, with the addition of appropriate inquiry. Equipment employed in the original Bruner-Goodman study (1947) was used for making the coin-size estimations. Unlike the procedure of Ashley, Harper, and Runyon, however, the coins were presented on the *S*'s left palm which he was permitted to hold beside the box. He was not permitted to remove the coin from his palm.

All *S*s used in this study had demonstrated their ability to manifest all of the usual deep trance phenomena including responsiveness to posthypnotic suggestions and the ability to experience what appeared to be total amnesia when this was suggested.

The procedure, briefly stated, was as follows: After the *S* was placed in trance, amnesia for his own life history was induced. He was then given a pseudo-life history which was essentially the same as that described by Ashley, Harper, and Runyon. The poor state was induced first, then the rich state, and finally the normal state. The *S* judged coin sizes in all three states. The same *S* was run with both imagined coins and with real coins presented in all three states. Also in all three states, he was given brass slugs which were called "lead," "silver," "gold," and "platinum." The brass was of a very whitish color so that it could conceivably have been the appropriate metal.

Results

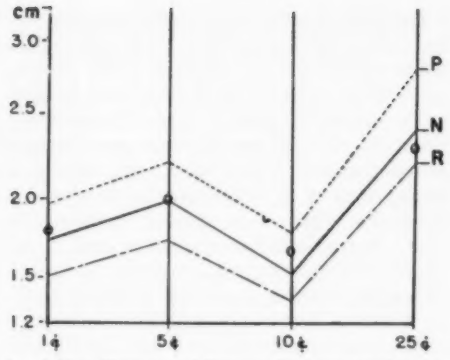
The results are summarized in Figure 1b, which presents the subjects' average coin-size estimates. The data are essentially identical to those obtained by Ashley, Harper, and Runyon (see Fig. 1a). The data on the size estimates of "slugs" successively called silver, gold, and platinum were also similar to those presented by Ashley, Harper, and Runyon in their series using a lead slug. All of the four subjects were able to describe correctly the purpose of the experiment and the hypotheses of the investigator who originally designed the experiment.

Discussion

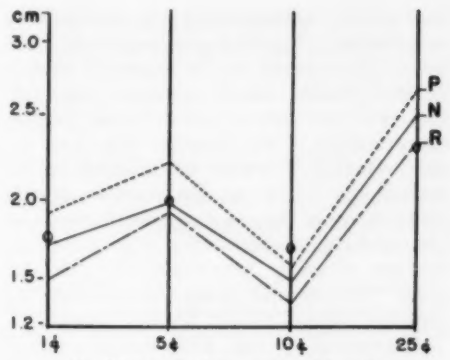
The data from the pilot study imply that the present procedure effectively reproduces that of Ashley, Harper, and Runyon. Both in terms of the quantitative results and the observed behavior of our *S*, no significant differences emerge.

The only essential difference between these data and those obtained by Ashley, Harper, and Runyon relates to the inquiry procedure. The results confirm the first two hypotheses. 1. The *S* in an experiment is able to express some demand characteristics of the procedure, if careful inquiry is done and his initial resistance is penetrated in a clinical fashion. 2. The majority of *S*s may perceive the same demand characteristics of the experiment and these may be the same as the hypothesis being tested. However, the third hypothesis has yet to be dealt with.

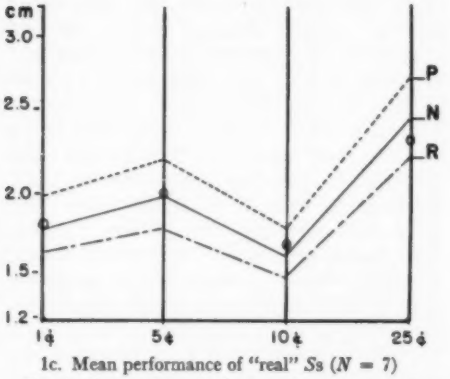
It is interesting to note that two of the four *S*s who were specifically questioned about this point denied vehemently that they were influenced during the experiment by an awareness of the experimenter's hypothesis. But the *S*'s verbalization during inquiry cannot be accepted at face value. As long as the *S* recognizes and is able to verbalize the demand characteristics of the experiment, they may play a significant role in his experimental behavior, although to demonstrate that they do so requires supporting evidence. It is with this further evidence that the main study is concerned.



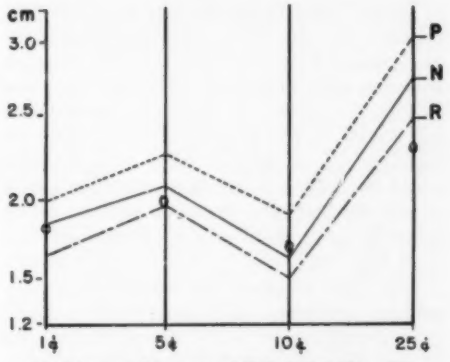
1a. Ashley, Harper, and Runyon data converted to cm



1b. Pilot study data: means of four *S*s



1c. Mean performance of "real" *S*s ($N = 7$)



1d. Mean performance of "fake" *S*s* ($N = 10$)

FIG. 1. COMPARISON OF ASHLEY, HARPER, AND RUNYON DATA WITH DATA FROM PRESENT REPLICATIONS (Legend: [R] = Rich; [P] = Poor; [N] = Normal; [O] = Actual coin size)

* One *S* (B. S.) was highly atypical and therefore excluded. See Fig. 2(b) for his performance.

Main Experiment

While the data found in the pilot study are consistent with the hypothesis that the demand characteristics of the experimental procedure may determine behavior, they are open to several serious objections.

The greatest single problem relates to the technique of inquiry and the interpretation of the data obtained in this fashion. It is important to have an objective method of rating how well the *S* perceives the demand characteristics of the experimental situation. The study was therefore designed so that the *S*'s inquiry would be rated by independent judges who did not have available to them the *S*'s data, but who would only have the opportunity of reading transcripts of the inquiry.

Another problem is a bias inherent in the inquiry procedure. Some *S*s who do not perceive the demand characteristics while engaged in the formal experimental procedure may perceive them during the inquiry. In such a case, and if the demand characteristics rather than the experimental variables determine the response, then the inquiry may indicate that the *S* should have responded a certain way when in fact he did not. However, the reverse should not occur.

The question still remains as to whether the *S*'s perception of the demand characteristics is responsible for his behavior, or whether it is due to the operation of the "intended" experimental variables. This question was dealt with by including a control group

that could not conceivably be construed as experiencing a "psychologically rich and poor state." If it could be demonstrated that a group of *S*s who do not experience the "rich and poor state" but are exposed to the demand characteristics of the procedure also show the data reported, it would be justifiable to attribute the results to the demand characteristics rather than to a presumed change in the psychological organization of the individual because of being "psychologically rich and poor." The control group thus permits inferences without reliance on the inquiry.

A group of *S*s who were not in hypnotic trance and did not manifest amnesia should provide such a control group. They would, of course, have to go through the same procedure as the "real trance" group. Such a group of *S*s would be asked to "play act" being in hypnosis and go through the whole procedure as if they were real *S*s. This group of *S*s would not truly consider themselves as psychologically rich or poor. In these *S*s no amnesia could be induced, and their behavior would clearly be that of a group of persons acting under three different sets of instructions--act as though you were poor, rich, and yourself.

This type of procedure is open to an important objection. Experimenter bias could play a major role. While the procedure and the wording of instructions would be the same, it would be possible unwittingly to include a variety of cues which could differentially shape the behavior of the two classes of *S*s. A blind technique is thus necessitated, in which the experimenter would not know which *S*s were "real" and which were "fake."

Such a stratagem presupposes that a "fake" *S* can simulate hypnosis sufficiently well to deceive the experimenter. However, there is a widely held opinion in the literature that it is impossible to simulate hypnosis successfully (Jenness, 1944; Stokvis, 1955). Cursory attempts by the author to have *S*s fake trance showed that the *S*'s efforts were half-hearted and obviously transparent.

In the usual faking situation, the experimenter knows that the *S* is faking, and the *S* is aware that the experimenter knows it; the usual purpose of this situation is to demonstrate the difficulties of fooling an experienced hypnotist. Clearly, the experimenter is not really expecting the *S* to be able to carry out his task effectively, and the *S* is aware of this. Thus, the *S*, who is anxious to please the experimenter, is in actuality motivated to give an unsuccessful performance. Furthermore, since the *S* is aware that the experimenter knows that he is acting, the *S* feels, with good cause, that it is impossible to deceive the hypnotist. There is a marked tendency to smile during induction procedure and in response to suggestions that might be construed as foolish, as well as to ask "How am I doing?" at intervals. Any suggestions that evoke even mild discomfort are followed only briefly and half-heartedly.

Most classical texts and modern authorities agree that hypnosis cannot be faked easily and "if a subject attempts to fake, tests for anaesthesia will permit ready recognition" (Estabrooks, 1948; LeCron & Bordeaux, 1947, p. 103; Mayer, 1951). However, the author has, upon two occasions, been taken in by *S*s who had apparently faked their way through the procedure and who subsequently disclosed the fact. In discussions with other hypnotists, he found that all who had had considerable experience could recall similar instances.[4] These experiences are usually explained by stating that such *S*s must really have been in the trance state or they would not have been able to act as well as they had, a view that is supported by the literature on hypnosis. The report of the *S* that he has not really been in hypnosis is thus lightly passed over, since "subjective reports are really not reliable." It is the author's opinion that it is dangerous to ignore the conviction of the *S*, expressed in good faith, that he did not experience the amnesia or anaesthesia or any other trance phenomena that he appeared to have experienced; and that it is indeed possible to construct a situation that would facilitate the successful "faking" of a hypnotic trance.

To do so it is necessary to motivate the *S* to

[4] In this respect, it is interesting to note that Pattie (1937) was sufficiently concerned with being deceived by a faking *S* that he asked each *S* to sign a statement "I, realizing that the experiment performed on me will probably be published in a scientific journal, solemnly declare that I was not faking or imitating the hypnotic trance but that I was genuinely hypnotized and do not remember the events of the experimental periods."

act the part of a hypnotic *S*—not to go into a trance—and for him to believe that it is possible to convince the hypnotist that he really is in trance. Accordingly, a situation was set up in which the *S* knew that the hypnotist would not know in advance whether or not he was really in trance. Further, it was so structured that the *S* would perceive the desire of the experimenters for him to do a successful job of faking. Finally, the situation was presented as a challenge that provided ample motivation for the *S* to tolerate unpleasant suggestions or tests.[5]

Training of Subjects

College students from Harvard and Brandeis were used as *S*s. All *S*s had been present during a lecture about hypnosis and a demonstration of the phenomenon given by the author in the context of an introductory psychology course. All *S*s were volunteers who agreed to take part in the study at the time of the lectures. No *S*s were paid. All had also been tested and rated by investigators other than the author for "suggestibility," in at least one screening session. All *S*s in the real trance group had further training with the author's associates. In order to satisfy criteria for depth of trance, they had to manifest the ability to respond to all the usual hypnotic suggestions with special emphasis on analgesia, visual hallucinations, negative auditory hallucinations, amnesia induced by suggestion, and simple posthypnotic suggestions.

The fake group of *S*s was drawn from the same population as the others but did not have special training in entering trance beyond the first screening session. They were all informed of their task in the manner described below.

The experimental procedure was administered to all the *S*s by the author, and he was given no information or cues by his associates as to which group any *S* might belong. Nor was he informed of the number of real and of fake *S*s. There were seven real *S*s and 11 fake *S*s.

Special Procedure for Faking Subjects

Prior to the experiment, the faking *S*s were told by an investigator other than the hypnotist:

> We're doing a study in which we would like you to take part. You recall some time ago you volunteered as a subject for some hypnotic experiments. We would like you not to go into a hypnotic trance; however, Dr. Orne, who will be working with you, will not know whether you are a real or a fake subject. He will go through trance induction with you, and then several tests and an experiment. We would like you to act as if you were entering a trance, but I don't want you to go into a trance. Your job is to fake the whole thing and to try to fool Dr. Orne. He knows that I may have some people trying to fake their way through the experiment but he has no idea who they are. I would like you, as a special favor, to try to do the best possible job of fooling him; to really convince him that you are in trance. This is quite difficult but a very important part of the experiment. I am trying to find out whether people are capable of doing this successfully. One of the tests which he will do with you will involve a harmless but somewhat painful electric shock. After the trance part, which you will have been faking, he will ask you some questions about it. Again, I don't want you to let on that you have been faking, and I want you to try to answer these questions as though you had been in hypnosis. Please keep on playing this part until I give you a signal. After the signal you can tell him what really happened. The signal will be that I will turn off the light on a box which will be on the table in front of you. I am sure that you will be able to do a good job of it.

The *S*s usually objected by saying that they did not know how a hypnotized *S* would act. They were told that although this feeling was understandable, others had been able to carry it off and that besides, they themselves had seen other *S*s go into a hypnotic trance during a previous demonstration.

The motivation for the *S* to act the role of the hypnotized individual is fairly clear. The author had been presented to the *S*s as an authority in the field of hypnosis and they had been taken into confidence by one of his associates in an attempt to fool him. Thus, on the one hand they were contributing to the advance of knowledge, on the other, they would have the pleasurable experience of "putting one over" on one of their teachers.

[5] After the completion of this study a previous reference to the use of a faking subject as a form of control was encountered. In an experiment on hypnotic pain suppression, Dynes (1932) asked a member of the psychology department to talk to *S*s privately and have them fake a trance the next time. The *S*s thought that he knew nothing of the plan and carried out the request to fake. However, Dynes himself did know and reports no difficulty in distinguishing this state from the real trance. It would have been interesting to know whether Dynes would have found it quite so easy to distinguish the behavior had he not known in advance.

Experimental Procedure

At the experimental session the *S* was exposed to trance induction. The Erickson hand levitation technique as described by Wolberg (1948) was used. After the *S* appeared to be in trance, a series of tests were done such as: hand clasp, analgesia, negative auditory hallucination, and positive visual hallucination.

When these tests had been completed, the *S* was told that his name would appear on his forehead and that he would feel it as though it were there in Braille. After this suggestion had been accepted, he was informed that if he would wipe this off he would forget everything about himself. Once he had accepted the suggestion of total amnesia, the rich state and the poor state was induced in a fashion identical to that used by Ashley et al. (1951). The *S* was then asked to make coin-size judgments of a penny, nickel, dime, and quarter, again using the original Bruner-Goodman box. He was required to make six coin-size judgments of each coin in each state. The final set of judgments was made in his normal state, but with the suggestion that he had no recollection of anything that had happened since he entered trance. The rich and poor states were randomized as well as the order of the coins. A balanced design could not be achieved in respect to coin order because of the number of *S*s used, but it was approximated as far as feasible. The hypnotist-experimenter did not know the key to the order and was merely told in each instance which state he was to run first and with which coin to start.

The procedure, used by Ashley, Harper, and Runyon and in the pilot study, of estimating the size of the lead slug which was called lead, silver, gold, and platinum, was omitted.

At the completion of the coin-size judgments, an inquiry was performed as previously described. This inquiry was the same for all *S*s. At the completion of the inquiry, both in the hypnotic and wake states, the light of the Bruner-Goodman apparatus was turned off, which was the signal for the fake group to tell what really happened. If the *S*s were real, this had no meaning to them, and the experiment was completed. Those of the fake group were asked a series of questions as to whether or not they had been in trance. They were also asked about their performance, why they had acted as they did on some points, and if and when they felt anxious. Finally, they were asked whether they thought that they had given themselves away; if so, when and why. All of these inquiries were completely tape recorded and transcribed so that, subsequently, judges could rate what each subject perceived as the hypothesis being tested.

Results

1. *Objective Findings–Group Results.* Figure 1 displays a comparison between the present data and those of Ashley, Harper, and Runyon (Fig. 1a); in Fig. 1c, the results of the hypnotic *S*s are averaged in the manner described by Ashley, Harper, and Runyon; in Fig. 1d, the results of the "fake" *S*s are averaged in the same manner. All three graphs present essentially the same configuration. In all cases the judgments in the poor state are the largest, judgments in the rich state smallest, while judgments in the "normal" state fall between.

2. *Results for Individual Subjects.* Figures 2a and 2b give the results for each "real" or "fake" *S*. *S*s varied widely in their response to the experimental task, not all of them yielding a configuration that corresponds to the group average. While Ashley, Harper, and Runyon unfortunately do not give their individual results, they report considerable variation.

3. *Comparison of Judges' Ratings with Objective Categorization.* Using analysis of variance for each individual *S*, it is possible to test statistically whether there are significant differences between *S*'s coin-size estimates in any combination of the three states and the direction of significant differences. Ignoring the "normal" judgments, the possibilities reduce to three categories: no significant differences between rich and poor, poor significantly larger than rich, and rich significantly larger than poor. Each *S*'s coin-size judgments were classified into one of the three categories on the basis of statistical analysis considering differences not significant at the .05 level as no difference.

The transcribed postexperimental inquiries were given to two independent judges to rate the *S*'s perception of the hypothesis being tested at the time of the experiment in terms of the same three categories. The judges had no contact with the *S*s or each other. Table 3 shows a comparison in terms of the

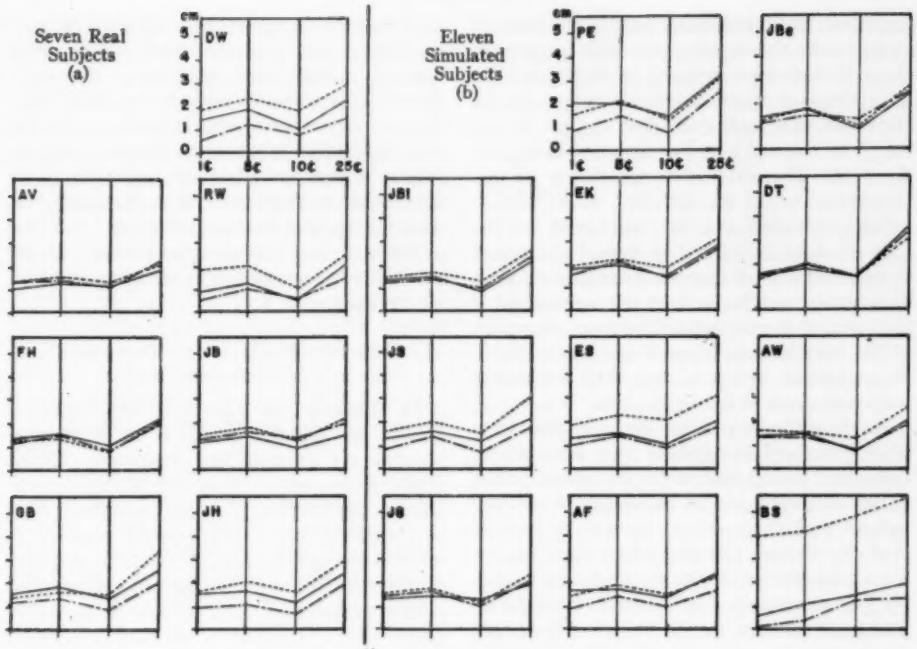


FIG. 2. GRAPHS OF INDIVIDUAL SUBJECTS
(Legend: [——] = Poor; [——] = Normal; [—·—] = Rich)

TABLE 3
A COMPARISON OF JUDGED AND ACTUAL CATEGORIES OF RESPONSE

Subject	Major Categories		
	J1	J2	Actual
G. B.	II	II	II
F. H.	II	II	I
J. H.	II	II	II
A. U.	II[a]	I	I
R. W.	II	II	II
D. W.	II	II	II
J. Ba.	I	I	I
P. E.	II	II	II
J. Bl.	II	II	II
J. G.	I	I	I
A. W.	II	II	II
E. S.	II	II	II
J. S.	II	II	II
D. T.	I	II	I
A. F.	II	II	II
E. K.	II	II	II
B. S.	II	II	II

NOTE.—Key to symbols: I. No significant difference; II. Poor significantly larger than rich; III. Rich significantly larger than poor.

Two judges were used (J1 and J2). Note that no *S* actually belongs in Category III and that neither judge placed any *S* within it.

[a] Judge undecided about I or II here but chose II as better estimate.

three categories between the ratings of the two judges and the individual's responses. There is a high degree of correspondence between the judges' ratings and *S*'s performance.

Testing the null hypothesis of no systematic correspondence between judges' rating and *S*'s coin-size judgments leads to its rejection for each judge (.015, Fisher Exact Test); combining the significance levels of the two judges leads to an over-all significance of less than .01.

Discussion

The data obtained from the seven hypnotized *S*s are essentially identical to the findings of Ashley, Harper, and Runyon, and virtually indistinguishable from the performance of the 11 stimulating *S*s. These results confirm the hypothesis that it is possible for an experimental design that omits a crucial aspect (hypnotic amnesia) of the original dependent variable to elicit similar responses as long as the same demand characteristics are present.

The subjective experience of members of the simulating group was radically different from that of the *S*s in deep trance. The *S*s readily

described their conscious efforts to "second guess" what the experimenter would expect of them if they were actually in hypnosis. The data obtained from the simulating group are the result of a concerted effort on the part of the *S*s to respond in a way identical to hypnotized *S*s. The subjective experience of the hypnotized group was different. While clinical inquiry revealed the *S*s' perception of the author's expectations, they denied that these factors had any effect upon their performance. This denial on the part of the hypnotized *S* does not, of course, mean that their perception of the experimental purpose was unimportant. It does mean, however, that they themselves were not aware of its significance.

An investigation of the demand characteristics perceived by each *S* may account for individual results that did not conform to the group average, as an examination of the judges' ratings confirms. It was discovered that the inquiry procedure had not been refined sufficiently to permit prediction of the *S*s' performance in the "normal" state. However, performance in the "rich" and "poor" states could be predicted with a high degree of accuracy from the judges' ratings of the *S*'s perception of the experimental purpose. No *S* reversed the expected trend by making his coin-size judgments larger in the rich state than in the poor state. No *S* was rated by either judge as having perceived this to be the hypothesis of the experiment. Twelve subjects made the coin-size estimates significantly larger in the poor state than in the rich state. All 12 *S*s were rated by both judges as having perceived this to be the author's hypothesis. Five *S*s failed to significantly differentiate their coin-size judgments between the rich and poor state. Of these five subjects, four were rated by either one or both judges as having failed to perceive the demand characteristics of the experiment.

The inquiry data thus support Hypothesis 3*a*, that if the majority of *S*s perceive the same demand characteristics, then *S*s who fail to perceive these demand characteristics should not show the behavior characteristic of the group.

The present experiments do not bear on the validity of the Bruner-Goodman effect. The Ashley, Harper, and Runyon experiment was used, rather, as an example of a study that appears methodologically sound, but in which demand characteristics seem to be the major determinant of the *S*'s performance. The implications seem clear: demand characteristics may determine behavior in hypnotic experiments. Before an effect can legitimately be attributed to hypnosis, it is necessary to demonstrate that it is not primarily a function of demand characteristics. Such proof appears to require the use of blind techniques and adequate inquiry.

THE INFLUENCE OF MOTIVATION ON HYPNOTIC BEHAVIOR[6]

In studying the nature of the hypnotic trance, the question arises as to which phenomena are primary and consistent components of the trance state and which are secondary derivatives. Let us postulate that increased motivation is a constant accompaniment of the hypnotic state. The present phase of the research was designed to show that certain phenomena long viewed as part and parcel of the hypnotic state may more parsimoniously be viewed as derivatives of increased motivation, and can be reproduced *pari passu* by other motivational techniques that have no direct relationship to hypnosis.

For years it has been claimed that there is an increase in physical capacity during the trance state. In part this claim has been based on casual observation, the favorite example being that of the stage hypnotist who places a subject in deep trance across two chairs and permits one or more individuals to stand or sit upon him. This "experiment," with variations, is often cited as irrefutable evidence for increased physical capacity. Another group of frequently cited observations are those concerning the ability of the subject to maintain his hand in an outstretched position for extended periods of time without evidence of fatigue. On the basis of this type of data, estimates of greatly increased physical capacity have been made (McDougall, 1926; Moll, 1904).

An early study by Nicholson (1920, p. 89) maintained that "during the hypnotic sleep the capacity for work seemed practically end-

[6] This experiment was originally reported in German (Orne, 1954).

less." Unfortunately, no quantitative data were given, and the study was poorly controlled. In a meticulous investigation, Williams (1930) showed no difference between hypnotic and wake states in the ability to maintain the arm in an outstretched position. However, this study failed to employ suggestions to the effect that the arm would not get tired and could not drop. In another similar investigation, using an ergograph and employing appropriate hypnotic suggestions, Williams (1929) found a 12 to 16% increment in the trance. More recently, Roush (1951) showed an increment in performance in hypnosis significant at better than the .05% level using the arm dynomometer, the hand dynomometer, and hanging by the hands, as measures of fatigue.

All the experiments performed by psychologists in the laboratory have followed orthodox scientific methods insofar as a standard set of instructions was given to the *S* to hold a weight, pull an ergograph, or perform a similar task in both the nonhypnotic and hypnotic states. The better experiments used the usual ABBA arrangement to control fatigue or practice effects. Any increment in performance was defined as an increase in capacity due to trance. It is necessary here to question the logic on which the interpretation of these results is based. While these experiments undoubtedly show that instructions given in trance state result in increased performance over that achieved by the same instructions in the wake state, they do not necessarily show an increase in capacity. Alternatively, the *S* may be more willing to exert himself while in hypnosis. The governing factor could be the increase in the *S*'s motivation to comply with the experimenter's request rather than an increased capacity to comply. The instructions, while identical in wording, may be experienced as quite different by the *S* in hypnosis and the waking state. The request to hold a weight at arm's length, given in trance, may be a highly motivating cue or "suggestion," especially if the *S* is told that he is to feel very powerful and not fatigued. The identically worded request in the wake state is perceived as a request by the experimenter and may be followed if good rapport exists between experimenter and *S*. However, as the discomfort of the task increases, the *S* becomes increasingly disinclined to comply. Viewed in this context, the reported experimental results do not necessarily imply that physical capacity is in fact increased in trance but, rather, that the trance state increases performance.

Procedure

Nine *S*s in deep trance were asked to hold a kilogram weight at arm's length. This was done in such a way as to derive maximal benefit from the peculiar nature of the trance state. Thus the *S* was told to hallucinate a table, and only after the table was both seen and felt by the *S* was the suggestion given that the right arm would feel no fatigue and no pain.

All the standard tests of deep trance were met in each *S*. A kilogram weight was placed in the *S*'s right hand, and the *S* was instructed to place it on the imagined table, to continue holding it with his fingers, and under no circumstances to drop it or his arm. Continuous suggestions were given to the effect that he would be able to hold onto the weight, that his arm would not get tired, etc., and that he would continue to see the table. The end point was when the *S* was no longer able to hold up his arm and began to come out of trance. At that point he was reassured, told to drop the weight, and deep trance suggestions were again given. After some minutes, and having made certain deep trance was again established, the *S* was awakened with a carefully induced posthypnotic amnesia. The *S* was not told the length of his performance.[7] For the second part of the experiment, which was done within half an hour of the first, the *S*, not now under hypnosis, was instructed as follows:

"This is a most important part of our experiment. It is very important for us to know your endurance and

[7] In the preceding section it was pointed out that the posthypnotic amnesia induced in hypnosis is not tantamount to an ablation of memory. One may be justified in assuming that the *S*s do not know their hypnotic performance, not because of the amnesia but, rather, because they were never informed of the length of time they held the weight in hypnosis. A common belief that the *S* in hypnosis has a perfect sense of time would lead to the conclusion that this is not an adequate safeguard. Fortunately, a very thorough study of the time sense under hypnosis was conducted by Guenther Klaus in a doctoral dissertation (University of Freiburg, Germany, 1948) which demonstrates unequivocally that the time sense is not improved by hypnosis.

TABLE 4
COMPARISON OF SUBJECTS' PERFORMANCE IN HYPNOTIC AND WAKE STATE

Subject	Hypnosis		Waking	
	Minutes	Seconds	Minutes	Seconds
1	4	05	5	33
2[a]	4	40	6	25
3	4	38	8	06
4 (*a*)[b]	6	05	3	29
(*b*)	5	50	10	02
5	7	07	7	57
6	10	05	16	00
7	4	52	5	49
8	5	20	5	32
9[c]	4	57	2 (*a*)	10
			5 (*b*)	09

[a] This experiment was performed in 1950. In 1957, it came to my attention that this *S* feels that he simulated completely throughout this experiment. At the time, I was totally unaware of this possibility and the *S* was in trance by all the usual criteria.

[b] *S* dropped the weight after 3′ 29″ in the wake state. The next day, care was taken to motivate him adequately. While the hypnotic performance was only 15″ below the previous day, his wake performance now exceeded 10′.

[c] This *S* suddenly dropped the weight without warning in the wake state after 2′ 10″. She was encouraged and after a 20′ time lag again held the weight. This time her performance was 5′ 09″. This performance in itself is better than her hypnotic performance of 4′ 57″; however, it might seem that the waking performance was better than this, as the 2′ 10″ period was not given credit.

physical capacity. What I want you to do is a very difficult task. It does not look difficult but it is. I want you to hold this kilogram weight at arm's length. Your hand will get tired and it will take great effort to do this. There is a natural tendency to drop the weight if your hand gets tired. However, it is vital that we get your true capacity. Surprisingly enough, our female subjects have been able to hold the weight for T minutes. [The time T given would be his previous performance during hypnosis rounded off to the nearest half minute.] Our male subjests have been able to hold the weight at least T + ½ minutes. I realize that this is a difficult and painful task. Just to make it interesting we will try a game. At T minus 2 minutes we will start you off at 5 cents. At T minus one and a half, we will double that and make it 10 cents. At T minus one, 20 cents. At T minus one half, 40 cents. At T, 80 cents and at T plus one half, $1.60.

Then the *S* was told that while we could not afford to pay over $1.60, we were, of course, interested in how long he could actually hold the weight. One final point was explained to him:

While we often feel that we are so tired that we cannot go on, this is not really true. One can rarely be so tired as not to be able to continue for 30 seconds. Accordingly, I would like you to give me one-half minute's notice before you actually drop the weight.

Results

Table 4 gives the results for the nine *S*s tested. All but one *S* in the wake state immediately exceeded hypnotic performance. This *S* held the weight for 6 min. 5 sec. in trance, a very remarkable performance, but in a subsequent wake state dropped the weight after only 3½ min. The exception demonstrates very clearly that it is necessary to ego-involve the person in the task and to convince him of his ability to do it. He reported that the seven minutes that had been given as an illustration of "average performance" had seemed so long, and his hand became so tired after three minutes that he felt convinced that he would be unable to come even close to the average, so therefore "why bother to try?" The next day the *S* was more carefully motivated and encouraged. He was then able to hold the weight for over 10 minutes.

Discussion

This experiment does not purport to prove that there is no increase in physical capacity in the trance state. Because of the motivating nature of the trance state, and the operational difficulty in obtaining equal motivational states, it becomes a technical impossibility to prove conclusively whether increased physical capacity is produced or not. The data, however, do show that the usually observed increase in performance of trance *S*s may be accounted for by motivational difference.

From a theoretical viewpoint the reinterpretations to which this study had led seem most significant. As long as we believe physical capacity to be in fact increased by the simple expedient of the induced trance, it becomes necessary to look for the focus of the trance in something neurophysiological. If, on the other hand, we can understand the apparent increase in physical capacity observed during the trance state in terms of differences of motivation, we are then led to view hypnosis in psychological terms. It is clear that this study says nothing about why the trance tends to increase motivation nor does it even prove that this is so. It merely shows that adequate

motivation in the wake state leads to levels of performance equal or better than those found in the trance.

An objection that might be raised takes the form of the question as to what would happen if similar motivational techniques were used in the trance state to those in the wake state. But this question has little bearing on the essential point. If application of these techniques should produce a trance performance greater than the wake performance, it could be interpreted as the result of combined effects of ego-motivation and the postulated increased motivation associated with hypnosis. If, on the other hand, performance in trance were not greater it could be argued that the type of ego-motivation used is not germane to the trance state.

It may, finally, be argued that the *S* in the wake state is, in fact, still in hypnosis, since the same experimenter who induced hypnosis conducts the second phase. Perhaps *S*s performed better in the wake state because of the demand characteristics of the experiment, i.e., my expectation that they should do so! It is not easy wholly to refute this argument. That all previous studies are open to the same criticism does not answer the question. The clinical observation that the *S* does not look, act, or feel in any way the same in the hypnotic part and the waking part appears much more relevant. Nevertheless, I hope sometime to repeat the study with the aid of another hypnotist who believes in "the power of hypnosis" and who, therefore, expects *S* to do better in hypnosis than in the wake state. If it were possible for me to enable *S*s subsequently to exceed their hypnotic performance, it would go far toward removing this objection, of which I was aware during the collection of data. A different way to check the results would compare the performances of "fake" and "real" *S*s, using the strategy developed in the section entitled Cues Implicit in an Experimental Design.

REAL VS. "FAKE" HYPNOTIC SUBJECTS

The "real-fake" technique, a method of enabling *S*s to simulate hypnosis, was developed to demonstrate the effect of role-play artifact on trance behavior. Differences between the real and faking *S* that cannot be accounted for by the faking situation may be viewed as characteristic of hypnosis. Some behavior shown by both groups may, of course, also be a true characteristic of hypnosis since the fact that someone is able to simulate a given type of behavior does not indicate that it is not genuine in the nonsimulating group. For example, that it is fairly easy to simulate the compulsive quality of the trance does not imply that this quality is not germane to hypnosis. However, assertions that volitional capacities can be transcended in hypnosis would seem to require showing that faking *S*s cannot produce similar performances.

The present use of real and faking *S*s in a blind design appears to offer several advantages. It permits a rigorous control, in terms of behavior, of inquiry procedures designed to elicit demand characteristics of experiments. In the faking situation, the variable assumed to be the cause of the behavior can be omitted. If such behavior still occurs, it can then be accounted for adequately by the implicit demands of the situation. In this respect the technique may have useful application in other areas of psychology. With respect to hypnosis itself, the technique permits a rigorous control of experiments that claim to demonstrate transcendence of volitional capacities. It also helps to eliminate many biases almost universally present in hypnosis research and throws into relief certain differences between the "real" and "faking" groups which can then be attributed to the hypnotic state. These differences may be highly germane to the essence of hypnosis and seem to have been obscured by the overwhelmingly impressive nature of the phenomenon.

The Faking Subject

The situation of the simulating *S* may be viewed as one where he can (*a*) contribute to research by pleasing one experimenter and (*b*) satisfy different other needs by fooling another experimenter. The possibility of simultaneously satisfying usually incompatible needs leads to a state of peculiarly high motivation. Thus, he is highly motivated to play the role of a hypnotized *S* and continues to do so even in the face of great discomfort, without, however, the alteration of consciousness and the subjective experience found in hypnosis.

It had been anticipated that there would be differences in the clinical appearance of the

two groups of *S*s, both when going into trance and while in trance. Since the hand levitation technique (Wolberg, 1948) was used to induce trance, it was expected that the "faking" *S*s would lift their hands in a manner clearly indicating active volition rather than permitting the hand to rise slowly and gradually in response to a subjective feeling of lightness. Similar differences were anticipated in respect to eye-closure. When the inability to bend the extended right arm was suggested, differences were expected in the *S*'s attempts to do so. The *S*'s responses to negative auditory hallucinations and positive visual hallucinations, analgesia and anesthesia, pseudo identities in the trance, and amnesia for the trance were all expected to be means of differentiating the two groups. The discovery that most of the simulating *S*s did not differ behaviorally from those in deep trance came therefore with considerable surprise. True, it was possible to distinguish most of the "faking" group from the "real" group, but this discrimination was based on a clinical diagnosis with an inordinate amount of subjective uncertainty and about 20% error. With prior knowledge that an *S* is simulating hypnosis it is very easy for the observer to discern cues that reveal the "faking" nature of the "trance" behavior. Without such prior knowledge, these cues are obscure and judgments can be made only with a great deal of subjective uncertainty.[8] If it seems obvious to the experimenter that the *S* is faking, systematic error would again be introduced, and could affect the results. In such cases it seems desirable to discard the subject from the experiment, as was done in two instances in the present study.

Response to Pain in "Fake" and "Real" Subjects

Traditionally, hypnotists have viewed the induction of analgesia for highly painful stimuli as the best test for clearly distinguishing simulating *S*s from the *S*s in deep trance. The author expected, therefore, to find that reactions to pain would distinguish between the "real" and "fake" groups.

The technique for producing pain for testing hypnotic analgesia which the author had been using routinely involves forcible flexion of the two terminal phalanges of the little finger and pressure of the experimenter's thumbnail on the dorsal aspect of the middle of the second phalanx of the *S*'s little finger. These techniques failed to discriminate between the two groups. The "faking" group reported pain but indicated that they had had little or no difficulty in suppressing their reaction.

In order to provide a more painful, though still safe, stimulus a Harvard inductorium was employed. Standard Grass silver electrodes were applied to the sides of the index fingers and fixed with adhesive tape. It was found that the "faking" group flinched less and tolerated more pain than the "real" group. Actually, almost all of the *S*s tolerated the maximal stimulation that could be applied by the inductorium.

Another procedure, not particularly painful in itself, was designed to maximize anxiety about pain. Negative auditory hallucinations were induced and at the same time an analgesia of one hand. A match was then lighted close to the *S*'s ear and held under the third finger of this extended "pain insensitive" hand long enough to produce painful heat without causing a second degree burn. The experimenter tested the time interval on himself each time in order to determine what constituted a safe but painful stimulus. In no instance was an *S* ever subjected to a blistering burn. Again it was found that there were no noticeable differences between the simulating group and the deeply hypnotized *S*s. However, postexperimental inquiry revealed that the simulating group experienced a high level of anxiety and some pain during the procedure, while those in trance felt little or no discomfort.

To summarize the results of these informal experiments with pain, no clear-cut behavioral differences between "real" and "fake" *S*s have as yet been found, no matter what type of stimulus was employed. However, on postexperimental inquiries a very real difference was found in the *S*'s attitude toward this part

[8] The experimenter is currently attempting to study systematically the cues perceived as indicative of faking by an observer who knows that the *S* is simulating to see if consistent patterns appear. However, results thus far indicate that observers who had experience with observing faking *S*s, knowing that they were simulating, did not thereby become more able to discriminate in the absence of such knowledge.

of the experiment and in his report of subjective pain.

Differences Between "Real" and "Fake" Subjects

The faking *S*'s condition is characterized by the presence of faking behavior and the absence of hypnosis. If we wish to regard true hypnotic behavior as the difference between what can be observed in faking and real *S*s, it is necessary to take into account the characteristics of the faking *S*s introduced by the act of simulating.

One of the observed differences is related to an attempt by the simulating group to avoid commitments whenever possible. In some aspects, the situation of the simulating group is analogous to that of the malingerer, although the motivation to malinger is different and usually stronger. The person who is simulating a disease attempts to give the appropriate behavioral and symptomological manifestations without subjectively experiencing the symptoms that would be present with the actual disease. The real hypnotic group is more akin to the hysteric who reacts to subjectively real symptoms without organic pathology. Rorschach studies on malingerers in the military service (Rosenberg & Feldberg, 1944; Ross, 1944) reveal a typical syndrome. The most significant aspect is a marked constriction of the record, with occasional bizarre responses, a feature that is understandable as an attempt to avoid giving oneself away in the face of an unknown situation while attempting to appear cooperative. Applied to the present situation, these features suggest a tendency for the simulating group to avoid responding whenever possible and when forced to respond in an unstructured situation to give the safest possible answer. Just this occurred. In the replication of the Ashley, Harper, and Runyon study, it was observed that the "faking" *S*s would tend to be more hesitant in giving a false life history after being told that they were rich or poor. This reluctance cannot be considered a particularly meaningful finding for the theory of hypnosis, since it should rather be attributed to the psychological situation analogous to malingering, and not to the presence or absence of hypnosis.

Differences observed which may be attributed to the presence or absence of trance were with respect to: (*a*) congruence, the appropriateness of a response to the apparent level of trance, and (*b*) trance logic, the apparently simultaneous perception and response to both hallucinations and reality without any apparent attempts to satisfy a need for logical consistency.

The "fake" *S*s tended to exhibit behavior that was incongruous with what clinically appeared to be the depth of trance. For example, an *S* who responded very slowly to hand levitation was given the suggestion that he could hear only my voice. This negative auditory hallucination was tested by rubbing the thumb against the first two fingers of the hand close to the *S*'s ear and asking him "do you hear anything but my voice?" A response congruent with the *S*'s behavior of slow hand levitation and in a medium trance would be either no response or a *slight* nod or shaking of the head. Frequently, a repetition of the question would be required to elicit even a nod of the head. An *immediate* and emphatic *verbal* denial of any auditory perception other than my voice was considered a typical incongruent response—incongruent both in respect to speed and modality of response.

The most reliable criterion, as it turned out, may be termed trance logic.[9] It refers to the ability of the *S* to mix freely his perceptions derived from reality with those that stem from his imagination and are perceived as hallucinations. These perceptions are fused in a manner that ignores everyday logic. A good example of this process is the reaction of many *S*s in hypnosis who see a hallucination of a person sitting in a chair and describe it as "This is very peculiar, I can see Joe sitting in the chair and I can see the chair through him." This type of reaction, when made spontaneously, was absolutely diagnostic of the "real" *S*. Unfortunately, it did not occur spontaneously in every *S*; it was found, however, that "fake" *S*s *never* gave such a response.

In order to obtain the same type of data

[9] This seems intimately related to primary process thinking or autistic thinking. To avoid the developmental, pathological, and theoretical implications of these concepts a new phrase seems preferable to describe the clinically observable phenomenon.

that had proved to be so characteristic of hypnosis, a special situation was devised. Throughout the experiments, one of the co-experimenters was in the room with the *S* and the author, taking notes. At the beginning of the experiment he sat to the side of the *S*, well within his visual field. Once the *S*'s eyes were closed, the co-experimenter rose very silently and walked behind the *S* out of the *S*'s visual field. Subsequently, the *S* was instructed to open his eyes while still in deep trance and a hallucination of the associate sitting in the chair was induced. This hallucination was facilitated by the fact that during the initial part of the experiment, just prior to the *S*'s closing his eyes, the associate was actually sitting in the chair. After the hallucination appeared to be fully accepted by the *S*, and he appeared to be responding to the hallucination as though it were actually the associate, he was instructed to turn around and look at the associate now behind him with the question, "Who is that behind you?" Almost invariably the "real" *S* would look at the associate, then quickly turn back to look at the empty chair, and back at the associate. After looking back and forth between the hallucinated and the "real" associate, the *S*s indicated verbally that they were perceiving two images of the same person.[10] When asked about this, they tended to give bland responses such as "mirrors" or a "trick." Though they seemed somewhat confused, they did not become seriously disturbed.

The "fake" group, in most cases, either refused to see anyone behind them, or claimed that they could not recognize the person. Occasionally, they admitted recognizing the associate behind them and then claimed that the hallucination had vanished. (This reaction, while unusual among the real group, is occasionally found, especially if the *S* did not fully accept the initial hallucination.) Of approximately 30 "faking" *S*s, only two acted as if they saw two images of the same individual. The others, when asked during post-experimental inquiry about the reason for their response, gave a very significant answer. They stated that I had instructed them to hallucinate, the associate in the chair and therefore, when I asked who the individual behind them was, they had to deny seeing him, or if they saw him, recognizing him. After all, there was only one such individual and I had already told them that they were supposed to see him sitting in the chair. This logical conclusion determined the response given by simulating *S*s; it did not occur to the overwhelming majority of the "real" *S*s who saw two images without any difficulty. This finding appears to represent a valid and significant difference. The "real" *S* responds to a subjectively real image of the associate in the chair. When asked about an objectively real image of the same individual, he is able to perceive this as well. He can respond to perceptions that are subjectively real and determined by the suggested environment, as well as to his actual perceptions of the real world, without attempting to satisfy a possible need to make them logically compatible. The absence of expression of a need for logical consistency seems, at this point, to be one of the major characteristics of hypnosis.

Because it is my belief that the "essence" of hypnosis will be found in the subjective experiences of the *S*, I have become increasingly interested in a series of techniques attempting to obtain data about the actual feelings and experiences of the *S*. In the future, I intend to develop inquiry procedures that will include "casual conversation" with another *S* who, in reality, is an investigator. It is hoped to elicit cooperation from the *S*'s friends. Casual preliminary attempts using such procedures indicate that material obtained in this way may be quite illuminating and not accessible to direct inquiry by the experimenter.

A POINT OF VIEW TOWARD HYPNOSIS

While much of the research described here appears to be explaining away the hypnotic phenomenon, the intention is rather to differentiate its valid and significant aspects from what might be termed artifact. One of the problems inherent in any study of hypnosis is that of its definition. There is high consensus of opinion about what constitutes hypnosis in terms of a variety of scales. However, the

[10] This situation was originally discussed by Milton H. Erickson in a personal communication.

essential characteristics have remained obscure. A great many investigators have become impressed and fascinated by the apparent transcendence of normal physiological capacities in hypnosis. The present research program has made me increasingly skeptical of the experimental data that purport to support this view. However, clinical data obtained both by others and myself seem to show in a dramatic way that responses can be evoked in some *S*s which they themselves could not perform voluntarily. Such phenomena seem to be limited to *S*s who have a peculiar disposition in this direction. For example, authenticated cases of hypnotically induced blistering have been achieved only in individuals with previous dermatological histories.

One might hypothesize that the capacity to produce marked physiological alterations in hypnosis is confined to persons who have a readiness to somatize in the organ system being investigated, which will usually have been demonstrated by a history of similar pathology occurring spontaneously. Such findings do not preclude the possibility, of course, that a transcendence of normal volitional capacities in some areas may eventually be established in the laboratory as unequivocally due to hypnosis.

Aside from the controversial issue of such changes in physiological capacities, it appears that a universal effect of hypnosis on any *S* in deep trance can be delineated in terms of his subjective experience. Experience, after all, is not to be taken as an ephemeral or unimportant aspect of hypnosis but, rather, as extremely significant and, to the *S*, dramatic and striking.

Any *S* who has experienced deep trance will unhesitatingly describe this state as basically different from his normal one. He may be unable to explicate this difference, but he will invariably be quite definite and certain about its presence. Thus, one of the characteristics of hypnosis is that in deep trance the *S* experiences the state as discontinuous from his normal waking experience (though not always in the intermediate stages of trance). Hypnotic trance differs from pathological states, which may also be discontinuous, in that the *S* enters and leaves the state in accordance with previously established "rules of the game."

Another aspect of this altered subjective state is one which the *S* describes as an inability to resist a cue given by the hypnotist. (Interestingly enough, if the *S* *before entering the trance* decides not to follow a specific suggestion he is able to resist it.)[11] The uniformity with which this compulsive quality is reported tends to make us accept it as a characteristic of hypnosis. However, it will not emerge as a difference in behavior between real and faking *S*s.

Finally, an important attribute of hypnosis is a potentiality for the *S* to experience as subjectively real suggested alterations in his environment that do not conform with reality. In trance, the waking distinction between an imagined idea and what is perceived externally to the organism fades, and images may be perceived as originating from external reality. Thus, the waking individual, no matter how hard he tries to imagine that he saw someone sitting opposite to him, might at best be able to evoke some kind of imagery but would always be aware of the distinction between this and reality. The *S* in deep hypnosis may well be unaware of the distinction, though at some level he will always be able to discriminate.

In sum, the principal features of the hypnotic state are seen as changes in the subjective experience which are characterized by (*a*) discontinuity from normal waking experience, (*b*) a compulsion to follow the cues given by the hypnotist, (*c*) a potentiality for experiencing, as subjectively real, distortions of perception, memory, or feeling based on "suggestions" by the hypnotist rather than on objective reality, (*d*) the ability to tolerate logical inconsistencies that would be disturbing to the individual in the wake state.

SUMMARY

This paper has attempted to delineate some aspects of hypnotic phenomena which can be

[11] However, suggestions that are inconsistent with the basic "rules of the game" governing the implicit contract between hypnotist and *S*, as seen by the *S*, are, as a rule, not followed: e.g., antisocial and self-destructive acts, or any other suggestions running counter to basic ego needs or superego inhibitions.

rigorously tested and established. The hypothesis that the subject's "knowledge" regarding behavior in hypnosis influences his own hypnotic behavior was supported by an experiment. Students were exposed to a demonstration and lecture on hypnosis in which catalepsy of the dominant hand was mentioned as a common feature of trance behavior. Five out of nine volunteers exhibited this phenomenon under hypnosis. No students in a control group, who were given a similar lecture and demonstration but with no mention of catalepsy, showed the phenomenon.

An experiment performed by Ashley, Harper, and Runyon, which depends on hypnotic amnesia to explain the results, was repeated with the inclusion of a control group of subjects. These were subjects who simulated hypnosis but who were otherwise exposed to the same experimental situation as the hypnotic subjects. The behavior of the simulating group was indistinguishable from that of the "real" group, and both were indistinguishable from the results of the original study. Some doubt is thus cast on an explanation of the results in terms of hypnotic amnesia, and support is lent to the hypothesis that the demand characteristics of the experimental procedure may be a significant determinant of subject behavior.

In another experiment it was found that motivated subjects in the wake state held a weight at arm's length for a longer period of time than they did while in the hypnotic state. This result casts doubt on the notion that enhanced physical capacity is a primary characteristic of the trance state.

Differences between "real" and "fake" subjects were investigated. The major difference appears to be a tolerance by the "real" subject of logical inconsistencies.

It was concluded that in the absence of objective indices of hypnosis the existence of trance may be considered a clinical diagnosis. Until an invariant index of hypnosis can be established, such a diagnosis must be confirmed by the subject's report of alterations in his experience, since the real focus of hypnosis appears to lie in the subjective experience of trance.

REFERENCES

ASHLEY, W. R., HARPER, R. S., & RUNYON, D. K. The perceived size of coins in normal and hypnotically induced economic states. *Amer. J. Psychol.*, 1951, **64**, 564–592.

BINET, A., & FÉRÉ, C. *Animal magnetism*. New York: Appleton, 1888. (Trans.)

BORING, E. J. *A history of experimental psychology*. (2nd ed.) New York & London: Appleton-Century-Crofts, 1950.

BRUNER, J. S., & GOODMAN, C. C. Value and need as organizing factor in perception. *J. abnorm. soc. Psychol.*, 1947, **42**, 33–44.

COUÉ, E. Self mastery through conscious autosuggestion. *N. Y. Amer. Library Serv.*, 1922. P. 83.

DUMAURIER, G. *Trilby*. New York: Harper, 1895.

DYNES, J. B. An experimental study in hypnotic anesthesia. *J. abnorm. soc. Psychol.*, 1932, **27**, 79–88.

ESTABROOKS, G. E. *Hypnotism*. New York: Dutton, 1948.

HULL, C. L. *Hypnosis and suggestibility*. New York & London: Appleton-Century, 1933.

JENNESS, A. Hypnotism. In J. McV. Hunt (Ed.), *Personality and behavior disorders*. New York: Ronald, 1944. P. 470.

LECRON, L. M., & BORDEAUX, J. *Hypnotism today*. New York: Grune & Stratton, 1947. P. 103.

MCDOUGALL, W. *Outline of abnormal psychology*. New York: Scribner's, 1926.

MANN, T. *Mario and the magician*. Lowe-Porter (Trans.) New York: Knopf, 1931.

MAYER, L. *Die Technik der Hypnose*. Munich: J. F. Lelumanns Verlag, 1951. Pp. 142–143.

MOLL, A. *Hypnotism*. (5th rev. ed.) New York: Scribner's, 1904.

NICHOLSON, N. C. Notes on muscular work during hypnosis. *Johns Hopkins Hosp. Bull.*, 1920, **31**, 89.

ORNE, M. T. The mechanism of hypnotic age regression: An experimental study. *J. abnorm. soc. Psychol.*, 1951, **46**, 213–225.

ORNE, M. T. Die Leistungsfähigkeit in Hypnose und im Wachzustand. *Psychol. Rdsch.*, 1954, **5**, 291–297.

PATTIE, F. A., JR. The genuineness of hypnotically produced anesthesia of the skin. *Amer. J. Psychol.*, 1937, **49**, 435–443.

PAVLOV, I. P. The identity of inhibition with sleep and hypnosis. *Sci. Monthly*, 1923, **17**, 603–608.

ROSENBERG, S. G., & FELDBERG, T. M. Rorschach characteristics of a group of malingerers. *Rorsch. Res. Exch.*, 1944, **8**, 141–158.

ROSS, Y. D. The use of the Rorschach method in the Canadian Army. *Rorsch. Res. Exch.*, 1944, **8**, 159–161.

ROUSH, E. S. Strength and endurance in the waking and hypnotic state. *J. appl. Physiol.*, 1951, **3**, 404–410.

SARBIN, T. R. Contributions to role taking theory: I. Hypnotic behavior. *Psychol. Rev.*, 1950, **57**, 255–270.

SCHILDER, P. *The nature of hypnosis*. Gerda Corvin (Trans.) New York: International Universities Press, 1956.

STOKVIS, B. *Hypnose in der Aerztlichen Praxis*. New York: S. Karger Basel, 1955.

STRICKLER, C. B. A quantitative study of post-hypnotic amnesia. *J. abnorm. soc. Psychol.*, 1929, **24**, 108–119.

WEITZENHOFFER, A. *Hypnotism*. New York: Wiley Press, 1953.

WELLS, W. R. Experiments in "waking hypnosis" for instructional purposes. *J. abnorm. soc. Psychol.*, 1923, **18**, 389–404.

WHITE, R. W. A preface to a theory of hypnotism. *J. abnorm. soc. Psychol.*, 1941, **36**, 477–506.

WILLIAMS G. W. The effect of hypnosis on muscular fatigue. *J. abnorm. soc. Psychol.*, 1929, **24**, 318–329.

WILLIAMS, G. W. A comparative study of voluntary and hypnotic catalepsy. *Amer. J. Psychol.*, 1930, **42**, 83–95.

WOLBERG, L. R. *Medical hypnosis*. Vol. I. New York: Grune & Stratton, 1948.

Received January 22, 1958.

DRIVE STRENGTH AND SYMPTOM MANIFESTATIONS[1]

J. R. WITTENBORN
Rutgers University
AND NATHAN S. KLINE
Rockland State Hospital

THIS report describes a study of the validity of a new procedure for inferring drive strength. The rationale for this procedure takes its origin in Hull's proposal that the reaction potential of a response may be regarded as a function of two intervening variables, drive strength and habit strength (Hull, 1952). Habit strength is the variable which accrues from learning, i.e., from the frequency with which a response has been reinforced, and reaction potential refers to the energetic aspects of the response, e.g., latency, duration, amplitude, etc. Hull proposed that if one could assume relatively small individual differences in the habit strength of the response, the individual differences in the reaction potential of a response could be due to individual differences in drive strength (Yamaguchi, 1951). If, for example, one were considering individual differences in the reaction potential of a response which is a highly practiced one for all members of the sample, one could infer that these individual differences would be primarily expressive of individual differences in drive strength.

In attempting to make practical use of this principle for inferring individual differences in drive strength, one must be prepared to assume that for the sample in question there are no important differences in the strengths of the habits involved. One must, moreover, be prepared to make distinctions in the quality of the responses in a way which represents differences in reaction potential (Hull, Felsinger, Gladstone, & Yamaguchi, 1947).

The rating scales which we now seek to validate were constructed by the application of this principle. We were interested in inferring differences between individuals with respect to the strength of the drives involved in a given situation when one resists the requests of one's superior, seeks the attention of a superior, or expresses pleasure. We assumed that the various acts which may be involved in these classes of behavior were sufficiently well learned that they entailed no major differences in habit strength and could, therefore, be primarily expressive of differences in drive strength. The next step required that we be able to distinguish between persons with respect to differences in the reaction potential of the various classes of responses under consideration. Since we wished to infer drive strength on the basis of behavior ratings, our use of the various approaches to differences in reaction potential was analogical and not directly specific (as is the custom in most laboratory studies). Nevertheless, the concept of reaction potential was interpreted in nine different ways:

1. *Threshold*, the intensity or the force of the stimulating situation which is required to elicit a response, is a familiar description for performance in both physiological and psychological laboratories.
2. *Latency* is the interval separating the probable eliciting stimulating situation and the appearance of the response in question.
3. The *amplitude* of a response usually refers to the force with which the response is expressed and directly implies energy manifestation per se. (It could also be construed to refer to the extensity of motion.)
4. The *speed* with which the response is executed is a quality which is commonly used in the psychological laboratory.
5. The *duration* of the response is a familiar quality for the description of a performance.
6. The *refractory* aspects of the behavior have to do with the probability of a given kind of performance occurring repeatedly as contrasted with the quality of behavior where the performance once elicited cannot be readily evoked again for some time. We surmised that the stronger the drive, the less likely a refractory phase.
7. *Stimulus generalization* is a quality of performance which repeatedly appears in the experimental literature of psychology and refers to the diversity of the situations which will elicit the performance in question. We expected that the stronger the drive for a given response, the greater the variety of stimuli which will elicit it.
8. A kindred concept is *response generalization*. This

[1] We gratefully acknowledge the assistance of John Blair of the Rockland Research Facility, and John Doebele, Ruth Raime, and Alfred Hall, who assisted in collecting and analyzing the data.

refers to the diversity of responses which may be elicited in a given stimulus situation. Our expectancy was that the stronger the drive for any one response, the greater the chance of similar responses.

9. The final quality has to do with *superfluity* of behavior, i.e., the diversity of responses which simultaneously occur in a given situation. In a given situation, some persons may express a diversity of response forms more or less simultaneously so that one would say that their behavior has a superfluous quality. In the same situation, another person might respond in one manner only.

It was possible to construct a rating scale for each of the nine interpretations of reaction potential. These nine rating scales were then applied to each of the three different classes of behavior under consideration: resisting superior, seeking the attention of superior, and expressing pleasure.[2]

In order to justify even a tentative interpretation that these rating scales, based on the various analogical interpretations of reaction potential, express individual differences in the strength of the drive involved in the behavior, it is necessary that the various rating scale interpretations for the reaction potential of a given class of behavior distinguish between patients in a similar way. The required pattern of intercorrelations was reported in an earlier publication (Wittenborn, 1957); in general, when resistance to the request of a superior was rated according to the nine different interpretations of reaction potential, the various rating scales tended to be consistently intercorrelated. The same pattern of consistent correlation was found for the scales applied to seeking the attention of a superior and to expressing pleasure.

Because of the rational considerations which determined the way in which these rating scales distinguish between individuals and because of the pattern of internal consistency among the rating scales, it was inferred that rating scales based on an analogical use of reaction potential *could* express individual differences in the strength of the drives contributing to the class of behavior under consideration. The present report is concerned with an effort to show that individual differences distinguishable on the basis of this kind of procedure *do* show a predictable relationship with certain independent indications of drive strength.

[2] Copies of the 36 rating scales are available upon request.

HYPOTHESES AND METHOD

It is commonly assumed that anxious patients have a low order of drive for the expression of pleasure. Depressed patients have a low order of drive for the expression of pleasure, and they are rarely motivated to initiate associations with others. It was further assumed that manic patients are characterized by drives to both resistive and attention-demanding behavior and that hysterics are likely to be particularly attention-demanding. Such implied behavioral differences among mental hospital patients can be summarized in quantitative form by the use of a set of symptom rating scales developed in the course of earlier investigations (Wittenborn, 1954).

The following hypotheses are stated for the purpose of examining the validity of the scales:

1. The various scales for inferring strength of drives for expressing pleasure would tend to be consistently related, but in an inverse manner, with the degree to which the patient's symptoms are characteristic of acute anxiety.
2. The various scales for inferring strength of drives for expressing pleasure would tend to be consistently related, but in an inverse manner, with the degree to which the patient's symptoms were depressive in nature.
3. The various scales for inferring strength of drives for seeking the attention of superiors would also be consistently related, but in an inverse manner, with the degree to which the patient's symptoms were depressive in nature.
4. The various scales for inferring strength of drives for seeking the attention of superiors would tend to be consistently related with the degree to which the patient manifested hysterical symptoms.
5. The various scales for inferring strength of drives for seeking the attention of superiors would also be consistently related with the degree to which the patient's symptoms were characteristic of the manic states.
6. The various rating scales for inferring the strength of the drives for expressing resistance to the requests of superiors would tend to be consistently related with the degree to which the patient's symptoms were characteristic of manic states.

For the purposes of the present study, the rating scales were applied to the behavior of 46 young patients at the Rockland State Hospital. These patients were carefully observed for a two-week period by student raters who were trained in the use of both the new scales

TABLE 1

CORRELATIONS* BETWEEN ACUTE ANXIETY CLUSTER SCORE AND REACTION POTENTIAL RATINGS

Ratings for Reaction Potential	Resistance to Requests	Expression of Pleasure	Seeking Attention
I Threshold	−.58	−.62	—
II Latency	—	.65	.65
III Amplitude	—	−.68	—
IV Rate	—	−.70	—
V Duration	—	—	—
VI Refractory	—	—	—
VII Stim. Gen.	−.79	−.68	—
VIII Response Gen.	−.55	−.79	—
IX Superfluity	−.52	−.78	—

* Significant at the .05 level.

Note.—These are tetrachoric *r*'s based on 46 young mental hospital patients.

TABLE 2

CORRELATIONS* BETWEEN DEPRESSED STATE CLUSTER SCORE AND REACTION POTENTIAL RATINGS

Ratings for Reaction Potential	Resistance to Requests	Expression of Pleasure	Seeking Attention
I Threshold	−.62	−.78	−.35
II Latency	.48	.82	.70
III Amplitude	−.32	−.62	−.59
IV Rate	—	−.75	—
V Duration	—	−.62	−.55
VI Refractory	—	.50	—
VII Stim. Gen.	—	−.75	−.57
VIII Response Gen.	−.35	−.82	−.48
IX Superfluity	—	−.80	−.48

* Significant at the .05 level.

Note.—These are tetrachoric *r*'s based on 46 young mental hospital patients.

TABLE 3

CORRELATIONS* BETWEEN MANIC STATE CLUSTER SCORE AND REACTION POTENTIAL RATINGS

Ratings for Reaction Potential	Resistance to Requests	Expression of Pleasure	Seeking Attention
I Threshold	.48	—	.42
II Latency	−.29	—	−.36
III Amplitude	.58	.28	.30
IV Rate	.54	—	.51
V Duration	—	—	.60
VI Refractory	−.68	—	—
VII Stim. Gen.	.52	—	—
VIII Response Gen.	.40	—	—
IX Superfluity	.34	—	—

* Significant at the .05 level.

Note.—These are tetrachoric *r*'s based on 46 young mental hospital patients.

TABLE 4

CORRELATIONS* BETWEEN CONVERSION HYSTERIA CLUSTER SCORE AND REACTION POTENTIAL RATINGS

Ratings for Reaction Potential	Resistance to Requests	Expression of Pleasure	Seeking Attention
I Threshold	.48	—	.62
II Latency	−.52	—	−.44
III Amplitude	.50	—	.46
IV Rate	.66	—	.61
V Duration	.44	—	.56
VI Refractory	−.70	—	−.58
VII Stim. Gen.	.52	—	.42
VIII Response Gen.	.46	—	.52
IX Superfluity	—	—	.52

* Significant at the .05 level.

Note.—These are tetrachoric *r*'s based on 46 young mental hospital patients.

developed for the purpose of inferring drive strength and in the use of the symptom scales. The ratings were made with unusual care, and the raters did not know the purpose of the ratings or the manner in which the data would be analyzed.[3]

RESULTS

As scrutiny of Tables 1, 2, 3, and 4 reveals, the data for our sample of 46 patients satisfied the requirements of the predictions. Our assumption that this new rating principle can yield distinctions which are expressive of differences in the strength of the patients' drives is not challenged by these data; as a matter of fact, the data suggest that such an approach permits us to infer differences in drive strength which may be plausibly related to other aspects of behavior expressive of differences in drive. Certainly this seems to be a definite possibility for inferring some of the drive differences implicit in symptomatic behavior.

It will be noted that the tables do not include all the possible correlations. Tetrachoric correlation coefficients were determined for only those relationships which were found on the basis of preliminary chi square analysis to be significant at the .05 level. Since tetrachoric correlations were used, and since the sample is not large, no formal comparisons are made between the various correlation coefficients. It is not suggested that any particular significance be ascribed to any one of the correlation coefficients, and, according to our theory, we have no a priori reason for assuming that any one of our estimates of reaction potential is either a particularly valid or a particularly questionable basis for inferring drive strength. For the purposes of discussion, we shall ascribe some practical

[3] The patients, adolescent and juvenile males, were observed for 6- to 8-hour periods each day.

significance where four or more of the estimates of drive strength are correlated with a symptom cluster.

Since the symptom cluster scores refer to familiar differences in patients and have some commonly accepted motivational implications, perhaps a discussion can best be organized about the symptom cluster scores. Table 1 shows, as predicted, that the most acutely anxious patients tend not to express pleasure. In addition to this predicted pattern of relationship, there is some indication that these patients are not impelled by strong drives to resist the requests of their superiors. In Table 2, we find that patients who are most depressed tend to be the least strongly motivated to express pleasure, and that these patients are also not strongly motivated to seek attention. In addition, there is some indication that these patients are not inclined to be resistive to their superiors. These trends are certainly plausible and are suggestive of the behavior patterns which are commonly observed to characterize the reactions of depressives. Thus, our new procedure reveals the similarity between depressed and acutely anxious patients: They are alike in the sense that drives toward the expression of pleasure are extremely weak and that drives to resist the requests of superiors tend also to be weak. Our procedure reveals a commonly noted difference between these two classes of patients; among the depressed patients, the drive to seek attention is distinctly weak, whereas among the anxious patients this drive seems neither characteristically weak nor strong. From such patterns of relationships it is apparent that there is a consistency between the strengths of the drives inferred from our procedure and the drive strengths implied by the patients' symptom manifestations.

In Table 3, we find as predicted that manic patients are characterized by drives to resist requests of superiors and by drives to seek the attention of superiors. Drives to express pleasure appear to be neither characteristically high nor characteristically low. It is interesting to observe in Table 4 that the drive pattern characteristic of manic patients is also characteristic of the patients who show the symptoms of conversion hysteria. The drives to seek attention are, of course, notoriously characteristic of hysterical patients, and this fact served as the basis for the predicted relationship confirmed in Table 4. The drive to resist the requests of superiors was not predicted to be characteristic of the hysterical patients. Nevertheless, such motivation is implicit in the hysterics' use of physical symptoms as an excuse for not playing their proper roles and as a way of showing resistance to others, particularly to their superiors.

DISCUSSION

It should be noted that we use the concept of drive in terms of the energizing qualities of behavior in the tradition of the learning theorists. Drive is not used here in the psychoanalytic sense, and our procedure does not contribute to an analysis of intrapsychic conflicts. According to a tradition beginning with Woodworth and characterizing many experimental psychologists, drive refers to differences in the energizing aspects of behavior. This is not to deny the value of analyses of intrapsychic dynamics; it is suggested, however, that the drive strengths inferred by the use of the present procedure are the drive qualities which are expressible in overt behavior and not the strength of covert drives which remain in mutually suppressive dynamic balance without finding overt expression. In other words, if one can make a distinction between overt *potentiated* drive and covert dynamically *bound* drive, our exploration is an attempt to infer the strength of potentiated drive expressed in overt behavior. This approach is not considered to be simply a description of the energetics of the behavior alone, because a distinction is recognized between the energetics, i.e., the reaction potential, of behavior which is well learned and the energetics of behavior which is not well learned. We propose that differences in drive strength be ascribed to differences in the reaction potential of equally well learned responses only. If one can not assume that the behavior involves responses equally well learned among the members of the sample, one must assume that individual differences in various interpretations of reaction potential express differences in the degree to which the responses are learned (habit strength) as well as possible differences in drive strength.

The writers see certain advantages in an approach which scales behavior on the basis of distinguishable individual differences in reaction potential. First, the distinctions in behavior are not made on the basis of the degree to which the behavior approaches some culturally determined norm, medical ideal, or ethical value system. The distinctions can be made for such relevant practical behavior, but they are organized on the basis of the energetics, i.e., the reaction potential, of the responses. It is feared that studies based on behavioral distinctions which are generated by practical values may have a significance which is more pertinent to the form of our society than it is to behavior in general. This leads to the second possible advantage of such a system of distinctions. During the '20s and '30s, most studies in behavior were conceived from the standpoint of basic science, using the physical and biological sciences as models and borrowing many of their traditions of measurement. During the '40s, however, studies in behavior were strongly influenced by the current problems of society and its members, and many of the new developments in assessment or measurement expressed distinctions in terms of these immediate practical problems. Another shift in emphasis is now discernible. With the introduction of energizing and tranquilizing drugs as aids in the treatment of behavioral problems there appears to be a resurgence of interest in the biological approach to the study of behavior. Accordingly, the writers believed that many research interests could be served by the development of concepts and procedures in measurement which were relevant to socially significant behavior, such as symptom formation, but which provided distinctions between individuals by the use of principles which were not dependent on distinctions in our social values, e.g., by the use of distinctions in behavior based on an analogical employment of essentially physiological variables.

In any new approach, considerable experience and conceptual refinement is required before we know either its limitations or how best to apply it. For some kinds of behavior the various indications of reaction potential may have a different expressive significance and may not, therefore, be similarly indicative of drive strength. Such possibilities must be explored before we can identify the situations in which the reaction potential principle may be properly applied to inferring strength of drive (Seward, 1955).

SUMMARY

This report indicates that a new principle for inferring the drive strength of human behavior has a practical relevance for external criteria. Estimates of the strength of drives among mental hospital patients for resisting the requests of superiors, seeking the attention of superiors, and expressing pleasure were found to bear an expected pattern of relationships with the symptom manifestations of these patients as gauged by the use of a standard set of psychiatric symptom rating scales. It is suggested that this principle for inferring drive strength may be usefully applied to researches concerned with psychopharmacological agents which have practical effects on behavior by mediating physiological changes.

REFERENCES

HULL, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.

HULL, C. L., FELSINGER, J. M., GLADSTONE, A. I., & YAMAGUCHI, H. G. A proposed quantification of habit strength. *Psychol. Rev.*, 1947, **54**, 237–254.

SEWARD, J. P. The constancy of the *I-V*: A critique of intervening variables. *Psychol. Rev.*, 1955, **62**, 155–168.

WITTENBORN, J. R. *Psychiatric rating scales.* New York: Psychological Corp., 1954.

WITTENBORN, J. R. Inferring the strength of drives. In M. R. Jones (Ed.), *The Nebraska symposium in motivation.* Vol. V. Lincoln: Univer. Nebraska Press, 1957. Pp. 191–259.

YAMAGUCHI, H. G. Drive (*D*) as a function of hours of hunger (*h*). *J. exp. Psychol.*, 1951, **42**, 108–117.

Received July 10, 1957.

VERBAL RATE, CONTENT, AND ADJUSTMENT RATINGS IN EXPERIMENTALLY STRUCTURED INTERVIEWS[1]

FREDERICK H. KANFER

Purdue University

A MAJOR obstacle in the examination of the psychotherapeutic process has been the problem of discovering those dimensions of analysis most relevant to behavioral changes as a function of therapeutic experience. Investigation of psychological and linguistic variables in interview behavior can broadly be divided into two clusters of approaches with some overlap in the methods usually employed. On the one hand, numerous studies have dealt with content analysis (Auld & Murray, 1955) or the categorization of information which the observer can extract from the subject's (*S*'s) verbal behavior. The choice of categories varies widely and is mainly dictated by the purpose of the analysis and the theory from which it is developed. The schemas range from semantic categories which require little inference on the part of the observer to those based almost completely on inferences about *S*'s emotional state or his dynamics. An example of the former is the use of the verb/adjective ratio (Balken & Masserman, 1940), whereas the latter is illustrated by Murray's (1938) need analysis or Bellak's (1954) interpretations of the TAT. The second group of approaches is represented by the study of structural speech characteristics. In this method, the dependent variable is a physical measure of speech behavior, such as rate or duration of utterances (Goldman-Eisler, 1954; Matarazzo, Saslow, & Matarazzo, 1956). The underlying theoretical basis rests on the argument that such differences in patterns of speech are related to other personality characteristics.

Those who argue for a structural approach to verbal behavior consider speech patterns to be stable personality characteristics and to be unaffected by mild changes in emotional attitudes connected with content (Chapple & Arensberg, 1940). Similarly, research on content variables has generally not attempted to deal with structural variables. It is quite reasonable to assume, however, that varied emotional states affect speech rate, for example, as well as content. Several divergent theoretical views have suggested the emotionality of the speaker as a major determinant of speech rate. Within Hull's framework (Benton, Hartman, & Sarason, 1955), anxiety is regarded as a drive state, and the level of verbal activity is related to manifest anxiety. On the basis of Skinner's approach, the properties attributed to anxiety also suggest changes in verbal output with increasing anxiety. In two earlier studies, Kanfer (1958a, 1958b) has shown that verbal rate increases as a function of shock-induced anxiety. In addition, clinical observers have repeatedly called attention to the patient's rate of speech as an index of his emotional state. In psychopathology, the direction of change in speech rate has been thought to depend on the patient's patterns of defenses. For example, blocking and increased verbal output may indicate a heightened emotional state, related to what is talked about.

The first step in integrating these approaches is to ascertain whether the content and structural dimensions show any correlation under varying emotional conditions. Although independent manipulation of these conditions is the eventual goal, a start may be made by identifying changes in content through controlling topics and by using the usual methods of content analysis to infer changes of the speaker's emotional state. Possible variations in structural variables may then be observed as a function of these changes.

The major purpose of the present study was to explore the relationship between verbal rate and rate variability and two content categories, topics and a content-inferred measure of anxiety. On the basis of the foregoing considerations, the following specific hypotheses were proposed:

Hypothesis 1. In a group of relatively homogeneous *S*s, interindividual similarities due to cultural influences

[1] This study was supported in part by Research Grant M-963 from the National Institute of Mental Health of the National Institutes of Health, Public Health Service. The data were gathered while the author was on the faculty of Washington University, St. Louis.

may exist with respect to the extent to which talk about certain topics modifies emotional states. If the topics evoke similar emotional responses due to cultural attitudes, then differences in verbal rate would be expected between topics.

Hypothesis 2. In addition to the effects of various topics on *all Ss*, idiosyncratic experiences of *Ss* in relation to these topics should result in arousing different levels of anxiety among them. In an interview setting, the extent of anxiety aroused in individual *Ss* by various topics is difficult to evaluate directly. However, anxiety level may be assessed from the content of the protocol. On the basis of earlier findings (Kanfer, 1958a), it is hypothesized that there is a positive correlation between verbal rate and a content-inferred measure of anxiety. Secondly, intra-individual rate differences among topics are expected to show the highest rates on those topics on which the individual's anxiety-ratings are also highest.

Hypothesis 3. In clinical work it has been suggested that such behavioral variability as restlessness, uneven speech, etc., are associated with anxiety. The third hypothesis therefore states that greater variability in speech behavior is related to such indices of high anxiety.

METHOD

Subjects and apparatus. Ss were 26 students of both sexes from the author's class in abnormal psychology. All were nonveterans, 19–21 years of age, and single. Data from two *Ss* had to be discarded because the highly personal material revealed forced *E* to modify the experimental procedure in order to reassure the *Ss* and to safeguard them from further stress. Illegibility of tape-recordings further reduced *N* to 20.

Interviews were conducted in the author's office, which was familiar to the students. The apparatus consisted mainly of recording and timing devices. A microphone was in full view of *S*. A beep signal was fed directly into the recorder at 5-sec. intervals. A hidden knife-switch controlled the onset of the beep signal. A modified desk-clock was used which could be reset at the start of the interview and could be read with an error of less than three seconds. The clock faced *E* and was half hidden behind books. Precautions were taken to avoid interruptions during the interview.

Procedure. Ss were told that the experiment concerned a study on a personality test. All *Ss* were given Rorschach tests by graduate students and were then given appointments "to answer several more questions." When seated for the interview, they were given the following instructions:

> This is a study in which we want to find out how well we can integrate our findings on a personality test and on a typical psychiatric interview of people who are relatively free of any mental illness. We will need as much information and as accurate an account as you can give us. Therefore, you can help us if you answer my questions by talking about everything which comes to mind when I raise a topic. Tell me about your attitudes and feelings and give whatever thought or example comes to mind. It is important that you omit nothing. You can talk freely, since this interview will remain confidential. I am going to record it since I cannot possibly write down all you say, but I will use a code number to identify the interview and no one will have access to any individual interview. After you have answered a few questions, I will discuss with you a little bit what the personality test indicated. Do you have any questions now?

Four warm-up questions about *S*'s school work and major interests were given in order to establish a pattern of continuous speech. With each succeeding question, *E* allowed longer time periods before asking another. When necessary, *E* explained that *S* had to continue talking about the topic and that *E* would not answer questions or ask any until each topic was exhausted. On the last question, three minutes were allowed. By that time, all *Ss* had learned that it was their responsibility to continue to talk, and all understood the instructions. After this warm-up period, the experimental session was begun.

The experimental interview consisted of five questions given to all *Ss* in the same order. After asking each question, *E* allowed three minutes of uninterrupted speech by *S*. The following questions were used:

1. Family Relationships (FR)
   Tell me a little about how you are generally getting along with your parents.
2. Self-Confidence (SC)
   Now I want you to tell me a little about yourself. Do you generally feel pretty sure of yourself in most situations?
3. Achievement (AC)
   Have there ever been times when you wished you were able to do something which you couldn't do or weren't competent to do?
4. Sex (SX)
   Do you generally feel that you are attractive to members of the opposite sex and that you get along well with them?
5. Emotional Maturity (EM)
   Do you think that you are emotionally grown up, that you are emotionally mature and able to accept things?

At the end of each three-minute period or as soon thereafter as *S* paused, *E* ended the

question by using one of two prepared summary statements which rephrased the question and indicated a qualified affirmative or negative reply to *S*. For example, after the first question (FR), *E* said: "In general, then, you feel that you get along pretty well (or, sometimes do not get along too well) with your parents."

E gave several practice interviews in order to learn to minimize such cues as nods or changes in facial expression. Eight interviews were given for the purpose of testing the equipment and procedure and standardizing *E*'s behavior. None of the pilot interviews were used in the actual study.

Treatment of data. Duplicate typescripts of all interviews were prepared. All records were transcribed verbatim, including such speech errors as repetitions, false starts, etc. Only vague sounds such as "eh" and "ah" were omitted. All punctuation was omitted. One copy was prepared for measuring speech rate and variability by marking the typescript at 5-sec. and 30-sec. intervals in accordance with the taped beep signal. Partial words, repetitions, and speech errors were counted as separate words. In view of the great variability of the 5-sec. units, all measures were obtained from 30-sec. samples.

For the evaluation of anxiety from content, a series of adjustment scales was constructed. *E* and two clinicians devised a separate five-point scale for each of the five topics.[2] The clinicians, who later rated all interviews, had no knowledge of the purpose of the study. Specifically, they were unaware that "structural measures" for the speech samples were available. The rating scales were anchored by supplying descriptive statements, specific to each topic, for each of the scale values. The raters agreed to deal with the interview content on a manifest level, i.e., their ratings were based only on such inferences which could be made directly from the data level and could be supported by specific portions of *S*'s content. Since most theoretical views hold that anxiety relates to adjustment and that good adjustment refers to a state of low anxiety which does not interfere with a person's effectiveness, a general view of the content was taken by instructing the judges to rate adequacy of adjustment. The concept of adjustment was essentially defined with respect to *S*'s social effectiveness. Thus, such manifestations as anxiety or defensive operations aimed at anxiety-reductions (including symptoms), lack of realistic, flexible, or tolerant self appraisal, and unrecognized conflicting behavior and attitudes or failure to attempt their resolution were considered indicative of poor adjustment. Ratings were assigned according to the frequency and intensity of such indices in the interview material. The rating scale was based on absolute criteria, so that no specific distribution of scores was expected for the population. After jointly rating two interviews from the pilot study, each judge independently rated the protocols. To reduce contamination between topics, each topic was rated for all *S*s in the order of topics, yielding one score for each *S* on each topic and a total adjustment score for each *S*, obtained by summing the five topic scores.

Two raters were used to obtain a measure of rater reliability. The rho of .75 between the two judges for the summed adjustment scores of each *S* indicated a reasonable reliability in judges' ratings of content. Rank-order correlations for individual topics ranged from .71 to .77. These coefficients indicate that judges' agreement did not vary greatly from topic to topic. Scores ranged along the entire scale. The mean adjustment ratings for topics varied from 2.6 to 2.8, indicating that no rater bias existed for any one topic. For statistical treatment, the ratings of the senior rater were used.[3]

RESULTS

Verbal rate. *S*s showed considerable variability and long pauses during the last-minute interval. In order to eliminate differences due to fatigue or lack of responses for a given topic, only the first four half-minute intervals were used in the analysis of verbal rate. Figure 1 presents the mean rates on all topics in half-minute periods. The analysis of variance summary in Table 1 indicates that the data do not confirm the hypothesis concerning the extent

[2] The writer wishes to acknowledge the contributions of G. J. Williams and R. A. Craddick, who served as raters.

[3] The writer is indebted to R. C. Bilger for his valuable suggestions concerning the statistical analysis of the data.

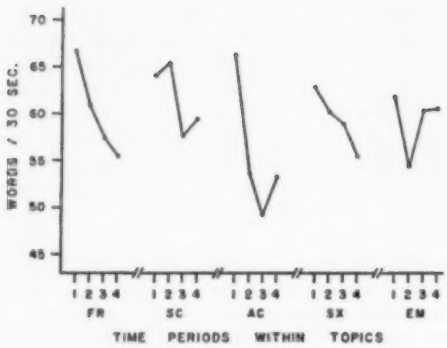


FIG. 1. VERBAL RATE WITHIN TOPICS IN 30-SEC. INTERVALS

of communality among *S*s with respect to topics. Even among the relatively homogeneous group used in this study, no group differences in speech rate were observed on the five topics. However, if the triple interaction term is considered an estimate of error variance, then differences in verbal rate due to *S*s' differential responsiveness to the various topics can be tested by T × *S*/P × T × *S*. As shown in Table 1, the resultant *F* is significant at $p < .01$. Therefore, some of the rate variance can be attributed to the different effects which the topics had on *individual Ss*. Over periods, speech rate decreased significantly, Fig. 1 showing the decrements by topic. As each new topic was begun, *S*s increased in their rate of talking. The *F* ratio for the P × T interaction shows no basis for inferring that the pattern of rate change within topics differed significantly among topics.

Adjustment ratings and verbal rate. Although the analysis of verbal rate indicated no differences between topics when groups were compared, it was hypothesized that the speech rate of individual *S*s is related to a measure of their anxiety level inferred from their replies to each topic. The following correlation coefficients between verbal rate and adjustment ratings were obtained: On FR, $r = +.467$; on SC, $r = -.099$; on AC, $r = -.070$; on SX, $r = +.376$, and on EM, $r = -.146$. High ratings represented poor adjustment. The *r*'s were significant at $p < .025$ and $p < .05$ on topics FR and SX, respectively, by one-tailed test. On these topics, high rate of speech was significantly related to poor adjustment. As a further test of the hypothesis, a sign test was run on the differences in rate between each *S*'s topics rated best and poorest in adjustment. Of the 20 *S*s, 15 showed a higher verbal rate on topics rated poorest in adjustment, yielding a *p* of .042. Thus, *S*s seem to talk significantly faster on those topics in which content suggests poorer adjustment than on those on which they are judged to be well adjusted.

TABLE 1
ANALYSIS OF VARIANCE OF VERBAL RATE DURING 30-SEC. PERIODS

Source	*df*	*MS*	*F*
Subjects (*S*)	19	2452.87	
Topics (T)	4	389.40	<1
T × *S*	76	420.39	1.75**
Periods	3	1284.15	4.01*
P × *S*	57	320.25	
P × T	12	239.41	<1
P × T × *S*	228	239.78	

* $p < .025$
** $p < .01$.

*S*s were then divided into two groups. All *S*s whose *total* adjustment ratings fell above the median constituted the "good adjustment" group. Those below the median constituted the "poor adjustment" group. An analysis of variance of the verbal rate data failed to show differences between the two groups. Furthermore, topics also yielded a nonsignificant *F* ratio. Thus, no additional information was gained by this analysis.

Variability of verbal rate and adjustment ratings. A variability measure was obtained by subtracting the lowest rate in each two-minute sample from *S*'s highest rate on the same topic. A rank-order correlation for variability scores, summed over all topics, and total adjustment ratings yielded a rho of .021. Thus, no relationship was found between general adjustment and variability of speech rate. However, correlations between variability and adjustment ratings *on each* topic yielded the following *r* values: FR, .396; SC, .389; AC, .263; SX, .207; and EM, .162. The values for FR and SC are expected to occur by chance less than once in 10 times. Thus, poor adjustment tended to relate to high variability on the initial two topics. As the interview proceeded, variability showed a decreasing relationship to adjustment.

A coefficient of concordance was run to test the consistency of variability from topic to topic. The resultant chi square of 33.92 is significant at the .02 level. Therefore, variability rankings within the population of *S*s used here show consistency from topic to topic. The results suggest, then, that the extent of variability of rate may be a fairly consistent individual characteristic.

DISCUSSION

The results show several general trends concerning verbal rate changes. First, rate declined considerably after *S*'s initial response to a topic. However, introduction of a new topic restored verbal rate approximately to the level originally found on the preceding topic. This consistent recovery suggests that no general adaptation occurred during the interview. If speech rate is related to anxiety, then the implications of these results suggest a rearousal of anxiety with each new topic over the interview.

Reflecting on the first hypothesis, one finds that the topics show no similar patterns of effect on rate for all *S*s. Thus, the assumption of a culturally determined, universal anxiety-arousing effect of the various topics is not borne out. When the adjustment ratings are used as criteria of anxiety and are derived from each *S*'s response to specific topics, the observation of significant rate increases from "best" to "poorest" topics within *S*s essentially confirms the second hypothesis. This finding is further supported by the significant $T \times S$ interaction, indicating that *S*s' verbal rates varied with topics, but that the topics differed in the emotional impact which they had on different *S*s.

With respect to the direction of change, the data show a facilitating effect of "anxiety" when this construct is related to adjustment ratings. This effect is in the same direction as observed earlier in a very different verbal situation (Kanfer, 1958a). It should be noted that *S*'s verbal productivity in both situations was relatively free of demands for high quality. The facilitating effect on quantity of verbal output does not preclude concurrent changes in quality (e.g., appropriateness with respect to a problem) which the present study did not attempt to measure. For example, increased redundancy and other content changes may occur. If they were found, they would suggest that emotional states may facilitate those aspects of verbal behavior which demand a lesser degree of organization while inhibiting more complex speech processes. Nevertheless, other factors may also contribute to variations in speech rate. The relatively low correlations obtained here preclude any use of the relationships for diagnostic purposes until a greater portion of the variance can be accounted for.

It is interesting to note that the highest correlations between rate and adjustment ratings were obtained on topics FR and SX. In comparison with the other topics, FR and SX differ in that they deal with areas in which verbal stereotypes or socially desirable "answers" are not readily available. In addition these topics required *S*'s evaluation of his own behavior in specifically defined interpersonal situations, whereas the others did not. Finally, FR and SX deal with attitudes in areas which are considered crucial by most personality theories. Similar results were obtained by Moore, Soderberg, and Powell (1952), who found that stuttering increased during the topic "parents." If stuttering is viewed as a manifestation of anxiety (Wischner, 1950), then Moore's results and the present ones would suggest that anxiety is more easily aroused in the parental and the sexual areas, and adjustmental differences become sufficiently dominant to affect such behaviors as verbal rate and stuttering.

Observation based on the content analysis revealed that most *S*s seemed to run out of relevant material on the remaining three topics and restructured the questions by relating specific incidents or elaborating some earlier statements. On FR and SX, however, *S*s continued their self-evaluation with respect to the topic and no such restructuring took place. Their tendency to stay with the topic may be reflected in the continued decline of rate during the last period.

In a study which derives an hypothesis similar to the present one from Hull's theory and uses Taylor MAS scores as indices of anxiety, Benton, Hartman, and Sarason (1955) suggested that their failure to obtain a faster rate in their high anxious group was due to *S*s' slowing down as their output increased. This

suggestion is verified by the present findings. The present results also point to Benton's use of a general criterion of anxiety as a possible source of failure to obtain rate differences between groups.

The data failed to establish a general relationship between speech variability and adjustment ratings. The higher correlations on the first two topics and their progressive decline suggest that poorly adjusted *Ss* are more variable in speech at first. Such adjustment-related differences did not persist over the interview.

The present study presented several problems in measuring rate. Verbal rate can be measured in any desired interval. However, the interval determines the specific definition of rate. A long interval masks such information as distribution and duration of pauses, "bursts" of verbal output, length of utterances, etc. On the other hand, a very short interval yields an extremely unreliable measure of rate which is strongly influenced by language structure. Although all these factors have been found to yield valuable information (Goldman-Eisler, 1954; Kelly & Steer, 1949), they cannot be investigated concurrently. The present measure was adopted after a preliminary study indicated that the 30-sec. interval yields relatively good stability for a measure of verbal units per time sample. The present rate measure can reveal changes in verbal output, but it does not reveal whether speed of utterance of a unit or pauses between units or other factors determine this change.

The data reveal the high sensitivity of the rate measure for individual response patterns in the experimental population. Thus, they confirm earlier research by Goldman-Eisler (1952) and others and support the contention that structural analysis of speech may be relevant to the investigation of personality. However, they also suggest the interrelationship of these physical measures with content-derived measures. The present study utilized only a very restricted range of adjustment ratings of a homogeneous population. The relationship between verbal rate and additional dimensions of adjustment needs to be explored further in populations representing a wider range of behavior patterns.

The adjustment measures admittedly represent an imperfect criterion variable for *S*'s "emotional state." They essentially reflect a clinical judgment criterion. Complete independence of the rate and adjustment measures cannot be proved despite attempts to avoid contamination of the adjustment ratings. Use of an external criterion (e.g., membership in a hospitalized or outpatient group or membership in a group whose emotional state is temporarily modified by experimental manipulations) remains a much more desirable step and is planned in further research.

SUMMARY

Twenty *Ss* from a selected, relatively homogeneous college population were given an experimental interview which they considered related to an earlier personality test. The interview was structured to allow collection of two-minute samples of uninterrupted speech, in reply to five questions relating to Family relationships (FR), Self-confidence (SC), Achievement (AC), Sexual attitudes (SX), and Emotional maturity (EM). Verbal output was tape recorded. Verbal rate was measured for 30-sec. intervals. Two clinicians rated the content of *S*'s reply to each question separately on a five-point scale devised for this study. Rate variability was measured by subtracting each *S*'s lowest rate from his highest rate on the same topic. There was no evidence that mean verbal rate differed among topics. However, speech rate decreased significantly within each topic. Furthermore, topics showed differential effects on the rate of *Ss*, as indicated by a significant topic × subject interaction. Significant correlations between high rate and poor adjustment ratings were obtained for topics FR and SX. In addition, it was found that *Ss* talked significantly faster on those topics on which their adjustment was rated poorest when compared with topics on which they were rated as well adjusted. High variability of verbal rate showed a moderate correlation, with poor adjustment on the initial two topics.

The results were related to a general hypothesis of the facilitating effect of anxiety on verbal rate, and the present findings were considered as suggestive of relationships between verbal rate, topics, and those aspects of *S*'s emotional state (anxiety) which overlap with a clinical evaluation of his adjustment derived from his verbalization.

REFERENCES

AULD, F., JR., & MURRAY, E. J. Content-analysis studies of psychotherapy. *Psychol. Bull.*, 1955, **52**, 377–395.

BALKEN, EVA R., & MASSERMAN, J. H. The language of phantasy: III. The language of the phantasies of patients with conversion hysteria, anxiety state and obsessive-compulsive neurosis. *J. Psychol.*, 1940, **10**, 75–86.

BELLAK, L. *The TAT and CAT in clinical use*. New York: Grune & Stratton, 1954.

BENTON, A. L., HARTMAN, C. H., & SARASON, I. G. Some relations between speech behavior and anxiety level. *J. abnorm. soc. Psychol.*, 1955, **51**, 295–297.

CHAPPLE, E. D., & ARENSBERG, C. M. Measuring human relations: An introduction to the study of the interaction of individuals. *Genet. Psychol. Monogr.*, 1940, **22**, 3–147.

GOLDMAN-EISLER, FRIEDA. Individual differences between interviewers and their effect on interviewees' conversational behavior. *J. Ment. Sci.* 1952, **98**, 660–671.

GOLDMAN-EISLER, FRIEDA. On the variability of the speed of talking and its relation to the length of utterances in conversations. *Brit. J. Psychol.*, 1954, **45**, 94–107.

KANFER, F. H. Effect of a warning signal preceding a noxious stimulus on verbal rate and heart rate. *J. exp. Psychol.*, 1958a, **55**, 73–80.

KANFER, F. H. Supplementary Report: Stability of a verbal rate change in experimental anxiety. *J. exp. Psychol.*, 1958b, **56**, 182.

KELLY, J. C., & STEER, M. D. Revised concept of rate. *J. Speech Dis.*, 1949, **14**, 222–226.

MATARAZZO, J. D., SASLOW, G., & MATARAZZO, RUTH. The interaction chronograph as an instrument for objective measurement of interaction patterns during interviews. *J. Psychol.*, 1956, **41**, 347–367.

MOORE, W. E., SODERBERG, S., & POWELL, DONNA. Relations of stuttering in spontaneous speech to speech content and verbal output. *J. Speech Dis.*, 1952, **17**, 371–376.

MURRAY, H. A. *Explorations in personality*. New York: Oxford, 1938.

TAYLOR, JANET A. Drive theory and manifest anxiety. *Psychol. Bull.*, 1956, **53**, 303–320.

WISCHNER, G. J. Stuttering behavior and learning: A preliminary theoretical formulation. *J. Speech Dis.*, 1950, **15**, 324–335.

Received December 18, 1957.

HOSTILITY IN RORSCHACH CONTENT AND OVERT AGGRESSIVE BEHAVIOR[1]

ALAN P. TOWBIN[2]

West Haven Veterans Administration Hospital and Yale University

IN VIEW of the evidence (Calden & Cohen, 1953; Gibby, 1951; Kimble, 1945; Klatskin, 1952; Lord, 1950; Sanders & Cleveland, 1953; Singer, Meltzoff, & Goldman, 1952) that Rorschach responses vary with situational factors, it seems reasonable to regard the test responses as a function of the testee's perception of the test situation, rather than as a function of his perception of the test stimuli alone. The general form of the relation between test responses and perceived situation is a matter of practical as well as theoretical interest, since it may be necessary to take these situational effects into account in interpreting the test data. Several writers (Joel, 1949; Miller, 1953; Sarason, 1950; Schachtel, 1945) have suggested that the test protocol may be regarded as a sample of behavior in an interpersonal situation, that is, behavior in a situation characterized by the mutual awareness and influence of the two participants. This conceptualization suggests the hypothesis that test responses are influenced by the perceived situation in the same way that the perceived situation influences behavior in other interpersonal settings. Furthermore, it suggests that given comparable sampling conditions, Rorschach responses and interpersonal behavior are positively correlated.

In the experiment reported here, the category of overt behavior studied was *aggression*, overt hostility directed at some object other than the person of the aggressor. It is a commonplace observation, as well as a theoretically important one (Dollard, 1947, pp. 269–281), that aggression tends to be inhibited as a function of the perceived capacity of the object of aggression to influence the condition of the aggressor. This capacity of the object is here designated *power*. *Status*, an individual's position in an institutional hierarchy, is closely though imperfectly correlated with his actual power over other members of the institution. Based on these considerations, the following predictions were made.

1. A positive correlation obtains between Rorschach test measures and independent behavioral measures of aggression.

2. The tendency of testees to produce aggressive Rorschach responses is inversely related to the perceived power of the examiner.

3. The tendency of testees to produce aggressive Rorschach responses is inversely related to the perceived status of the examiner.

METHOD

Design

The criterion measure of overt aggression was a history of such activity over a four-month observation period. *S*'s assignment to an assaultive (A) or a nonassaultive (NA) group was based on the liklihood of his manifesting aggression. *S*s in these two groups (A and NA) of schizophrenic patients were assigned at random to one of five conditions of Rorschach examination. Four conditions were procedures for influencing *S*s' perception of *E*'s power and status. The fifth condition followed routine hospital testing procedure and served as a control for the effects of the four experimental conditions. Rorschach records were scored for several categories of hostile content, only one of which, aggressive content (H_1), is relevant to the hypotheses tested here.

Subjects

Ninety-six male patients, diagnosed schizophrenic reaction (Committee on Nomenclature and Statistics of the American Psychiatric Association, 1952), were selected from the population of a large mental hospital. All were of at least average intelligence, under 50 years of age, and with one exception, had no diagnosed organic brain syndrome. The NA group was matched to the A group with respect to age, race, years of school attendance, diagnosis, interval since last Rorschach examination, and years of illness.[3]

[1] Based on a dissertation submitted in partial fulfillment of the requirements for the degree of doctor of philosophy in Yale University. The writer is indebted to and here thanks Seymour Sarason, Irving Janis, and Harold Kelley for their guidance in the execution of this study.

[2] This research was carried out while the writer was at F. D. Roosevelt Memorial V.A. Hospital, Montrose, N. Y. The writer wishes to acknowledge the support and the help of the psychiatric and psychological staff and psychology trainees who participated in this study.

[3] A detailed description of the selection, matching, and experimental procedures and of the scoring systems used may be found elsewhere (Towbin, 1955).

The A group ($N = 48$) was composed of patients who, over a four-month period, had two or more assaultive episodes recorded in the daily reports of the ward nurse. No NA patient appeared on the daily report during this period for either assaultive or self-destructive acts. To qualify as an assault, the behavior had to be of such severity that another person was struck or hospital property damaged, or it was reported that forcible restraint of the angry patient was thought necessary to prevent such outcomes. Thus, the group assignment of *S* defined his status with respect to a tendency to aggression in his overt behavior.

Procedure

Immediately before testing, *E* and a confederate (*C*) carried on a brief discussion in *S*'s presence, the purpose of which was to influence *S*'s perception of *E*'s power and status. From the patient's point of view, the most powerful person in the hospital environment is probably the ward chief, a psychiatrist or psychologist who daily makes administrative and therapeutic decisions on matters of importance to patients. In the High Power condition, *C* was the patient's ward chief. In his talk with *E*, *C* indicated that his decisions about patients were influenced by *E*'s test reports. The patient was thereby given reason to believe that what he said during testing session could influence decisions affecting his interests. In the Low Power condition, *C*s were trainees in the Clinical Psychology Service of the hospital and were strangers to the patients. The Low Power interaction between *E* and *C* in the presence of *S* contained reference neither to the uses of test reports nor to *E*'s influence.

Higher status is presumably associated with greater power over those in lower statuses in the same section of the hierarchy. High rank in one section of the hierarchy does not necessarily carry with it power over those in lower ranks in other sections. Since the symbols of high status are often similar from one section to another, e.g., terms of deference, a high status person may elicit inhibition of aggression in an individual in the absence of indications that he has real power over the individual. In the interactions between *E* and *C*, many cues to their status relation are conveyed in the High and Low Status conditions.

The interactions between *E* and *C* were as follows:

1. *High Power, High Status:* The power figure (*C*) conveys to *S* that *E* is a colleague and that *E*'s opinions, based on test results, contribute to his (*C*'s) decisions.

2. *High Power, Low Status:* The power figure (*C*) presents *E* to *S* as his assistant; it is implicit in the interaction between *C* and *E* that *E* is doing testing at *C*'s direction and reporting to him.

3. *High Status, Low Power: C* presents himself as *E*'s assistant; the interaction between *E* and *C* follows the course of the second condition with the statuses of *E* and *C* reversed and with neither having power in relation to the patient.

4. *Low Status, Low Power: C* interacts with *E* to convey to *S* that *E* and *C* are equals, neither of whom has power in relation to *S*.

5. *Control:* No confederate was used and the procedure usually employed at the hospital in introducing patients to the examination situation was followed.

The scoring system of Elizur (1949) and the results of other investigators (Gorlow, Zimet, & Fine, 1952; Sanders & Cleveland, 1953; Walker, 1951) using that scoring system were helpful in the development of the scoring procedure used here. An *aggressive Rorschach response* (H_1) was one in which hostility was directed from the testee to the card, whether the explicit object of the attack was a percept, the blot, or the test as a whole. H_1 was scored for attacks on the test ("This is crazier than I am!"), derogations of the blot (weird, messy), attribution of derogatory qualities to percepts (ugly, servile), blaming the card for *S*'s failure to respond, and pointing out faults or inadequacies in the blot or percept.

A second category called H_2 involved no activity. It was simply scored for objects with a customary aggressive function, e.g., gun, teeth, etc. Finally, hostility from card to testee (aggressive figures like witches or giants), from an unspecified source to the blot (percepts described as the objects of attack, as injured, etc.), and hostility within the card (two figures duelling), were all scored H_3. All responses received equal weight. *S*'s H_1 score was simply the number of Rorschach responses categorized H_1.

Twenty Rorschach records were drawn at random from the 96 cases, and one copy of each was rescored independently by two clinical psychologists in addition to *E*.[4] Since H_2 scores occurred relatively infrequently, they were combined with H_3 scores in computing the correlations between the three sets of scorings. The correlation coefficients for H_1 were .82, .95, and .72; the mean coefficient was .86. The coefficients for the combined H_2 and H_3 scores were .75, .89, and .81; the mean coefficient was .83. Thus, the scoring procedure showed an acceptable degree of reliability.

After the examination, *S* was asked his idea of the reason for the examination. *S* was considered to regard the test as consequential if his response mentioned an idea that the testing had to do with treatment, with evaluation of positive or negative personal qualities, with a decision affecting *S*, or if he said he thought someone had directed *E* to test him.

The interview records of two cases from each of the ten cells of the design were drawn at random, shuffled, and combined. These records were used to devise a scoring procedure likely to encompass the range of responses given by *S*s, but were not used for any item selection. These 20 cases were then returned to the remaining pool of protocols, and all interview records were scored in the same way. Twenty interview records drawn at random from the 96 were turned over to two clinical psychologists for rescoring.[5] Pairing the categorizations made by the investigator and the independent scorers gave 17 out of 20 possible agreements. Thus, the reliability of the categorization seemed satisfactory.

Only the first three responses to each Rorschach card were scored in order to control the variance in R.

[4] The writer wishes to thank Harold Wilensky and Jerome Singer for participating in this part of the study.

[5] Pearl Greenberg and Renata Calabresi carried out this rescoring, for which the writer here thanks them.

TABLE 1
SUMMARY OF ANALYSIS OF COVARIANCE:
$X = R$, $Y = \log (H_1 + 1)$

Source of Variance	X			Sum xy	Y			Corrected Y		
	df	ms	F		df	ms	F	df	ms	F
Treatments	4	10.7	—	1.27	4	.055	—	3	.046	—
Criterion	1	86.3	2.91	13.04	1	2.18	27*	1	1.90	26*
Interaction	4	7.8	—	1.55	4	.023	—	3	.022	—
Error	86	29.7	—	36.31	86	.080	—	85	.073	—
Total	95			50.72	95			92		

* Significance of .01 or above.

To estimate the degree of association between R, and the various measures of hostility (the criterion, H_1 scores, and H_2-plus-H_3 scores), the relevant distributions of scores were tabulated in two-by-two tables and chi square calculated in each instance. R was uncorrelated with the criterion of membership in the A or NA groups but correlated with H_1 and H_2-plus-H_3 scores to a degree significant beyond the .01 level. Consequently, in the tests of the experimental effects, the variance attributable to R was partialed out by means of the covariance technique.

The predictions stated earlier may now be put in more specific form:

1. The production of H_1 responses is positively correlated with membership in the A group.

2. *Ss* examined under the High Power condition produce fewer H_1 responses than testees examined under other conditions.

3. *Ss* examined under the High Status condition produce fewer H_1 responses than testees examined under other conditions.

RESULTS

Category H_1 includes both percepts reported by the testee and his remarks and comments about the test, the card, etc. In order to test Prediction I with respect to both reported percepts and remarks, separate analyses were carried out for each kind of behavior scored H_1 and for the two combined. The distributions of these scores were dichotomized and entered in a fourfold frequency distribution against group membership. The significance of the association was estimated by means of the chi square technique and phi coefficients were calculated as an estimate of the degree of relation. For H_1 over all, chi square was 26.33; phi was .52. For *percepts* scored H_1, a chi square of 24.14 and a phi of .50 were obtained. For *remarks*, chi square was 6.74; phi was .26. The relationships were all significant at the .01 level or beyond. The relation between the two subcategories of H_1 was significant beyond the .01 level; chi square was 9.09 and phi .31.

These results indicate that it is possible to identify behavior in the Rorschach protocol and in an extra-test sample which can be ordered to the same class of events. The separate analyses indicate that the relation between H_1 and the criterion is not attributable solely to the testee's admittedly social behavior, i.e., remarks addressed to *E*: the customarily scored "percept" is also positively correlated with the criterion. The significant correlation between the two subcategories supports the procedure of combining them in H_1.

In the test of the effect of experimental treatments on the H_1 response (Predictions II and III), the logarithmic transformation was applied to the H_1 scores. Bartlett's test of homogeneity of variance yielded a chi square of 14.224, with a confidence level of .11. Hence, analysis of covariance was applied to these data. The results of the over-all analysis of covariance are presented in Table 1.

None of the results of the over-all analysis of covariance were significant except the criterion group difference, which was significant beyond the .01 level. This is no more than another description of the high correlation between H_1 and the criterion.

Only the difference between the High Power and the other treatments is pertinent to Prediction II. The adjusted mean (Cochrane & Cox, 1950, p. 75) of the High Power group was .417; that of the Low Power-plus-Control groups was .503. The difference failed to reach an acceptable level of statistical significance ($t = 1.52$, $p = .064$), and the null hypothesis cannot be rejected.

The difference between the adjusted means of the High Status group and that of the Low Status-plus-Control groups was also not statistically significant. The t was .40, p .34. Thus, there is no evidence here for an effect of the status variable (Prediction III).

In addition to Category H_1, two other types of hostile content were scored. Although no predictions were made concerning these categories, it was supposed that a differentiation of types of hostile content in the Rorschach protocol could be made corresponding to a differentiation of types in extra-test interpersonal behavior. If correlations which were

comparable to that between H_1 and the criterion were obtained between H_2 or H_3 and the criterion of overt aggression, serious doubt would be cast upon the conceptual validity of such differentiation.

Category H_3 represents all types of hostile content in which there is a connotation of hostile activity, but it is other than from testee to card. The correlation of this score with the criterion was not significantly different from zero, nor was the correlation of any of the subcategories of H_3 with group membership. In an over-all analysis of covariance of H_3 scores, none of the effects approached statistical significance, *including* the criterion group difference.

Category H_2 was scored for objects with a customary aggressive function. The rationale for scoring suggests that H_2 falls somewhere between the aggressive Rorschach response (H_1) and the kinds of hostility lumped together in H_3. Perhaps it may be considered a potential or inhibited aggressiveness. The correlation between H_2 and the criterion fell short of statistical significance (phi was .19, p .07). However, in the over-all analysis of covariance of the H_2 score, the variance attributable to the criterion group difference was significant beyond the .01 level (F was 7.22, with df of 1 and 85). None of the other effects reached an acceptable level of statistical significance. For treatments, an F ratio of 1.89 was obtained; for interaction of treatments and the criterion group difference, the F ratio was 2.39.

It was hypothesized that testees examined under High Power conditions regard the test as having greater importance than testees in the Low Power and Control conditions. This assertion was tested by comparing the proportion of cases in the former group with the proportion in the latter groups reporting the examination as important when interviewed afterwards.

The number of *Ss* in the comparisons reported below was 78, because 18 cases either failed to respond to the question or responded too incoherently to permit scoring. The data are presented in Table 2.

The tendency to view the testing as important was positively correlated with being tested under the High Power condition (chi square was 3.55, p .032). There is no evidence of a relation between criterion of aggression and the attribution of importance to the testing (chi square was .32; p was $>.50$).

TABLE 2

PROPORTION OF *Ss* REGARDING THE TEST AS "IMPORTANT" AS A FUNCTION OF CRITERION GROUP AND CONDITION OF EXAMINATION

Examination condition	Criterion Group		
	Assaultive	Non-assaultive	
High power	.69	.81	.75
Other	.62	.40	.53
	.64	.58	

Inspection of Table 2 suggested that there was an interaction effect of the marginal factors such that the differential attribution of importance to the testing is associated with the condition of testing in the NA but not in the A group. A test of this second order difference yielded a critical ratio of 1.91, which has a significance level of .056. This result suggests that the experimental conditions had an over-all effect on attitudes only because of a differential effect on the NA group.

DISCUSSION

The positive correlation between the H_1 scores and overt aggressive behavior lends support to the view that the Rorschach protocol is a sample of behavior in an interpersonal situation. The power and status variables failed to influence H_1 scores to a statistically significant degree.

In connection with this result, it is worth noting that the assumption was made, though only implicitly, that the aggressive behavior of the testees in extra-test interpersonal situations is influenced by the power and status of those around them. This has not been documented, and there is some reason to believe that for many of the testees, this relation does *not* obtain. That is, assaultive patients may be precisely those who are minimally influenced by these factors in their everyday behavior. The interview data also suggest that this factor operated to weight the experiment against the hypothesis.

The fact that the various categories of hos-

tility (H_1, H_2, and H_3) showed different relationships to the criterion suggests that the differentiation of types of hostility in the Rorschach is meaningful, and it provides a basis for clarifying the behavioral implications of content in Rorschach interpretation.

SUMMARY

The Rorschach was administered to 96 hospitalized psychotic veterans, 48 assaultive and 48 nonassaultive, under one of five conditions. Test records were scored for aggressive content (H_1) as well as two other types of hostile content. It was predicted that there is a positive correlation between the behavioral and test measures of aggression, and that the test measure is influenced by the experimentally varied power and status of the examiner. A significant positive correlation between the H_1 scores and the behavioral measure of aggression was found. The effect of power on the H_1 scores, while in the predicted direction, fell short of statistical significance. There is no evidence of an effect of examiner status on the H_1 score. Two other categories of hostile Rorschach response were studied: H_2, scored for instruments with a customary aggressive function, correlated positively with the criterion (behavioral aggression). The other category of hostile response, H_3, scored for responses connoting hostile activity, but not directed from testee to card, showed no relation either to the criterion or to the experimental variables. The differentiation of types of hostile content, based on an analogy to social behavior, seems to be a meaningful and potentially fruitful procedure.

REFERENCES

CALDEN, G., & COHEN, L. The relationship of ego-involvement and test definition to Rorschach test performance. *J. proj. Tech.*, 1953, **17**, 300–311.

COCHRANE, W., & COX, GERTRUDE. *Experimental designs*. New York: Wiley, 1950.

Committee on Nomenclature and Statistics of the American Psychiatric Association. *Mental disorders*. Washington: American Psychiatric Association Mental Hospital Service, 1952.

DOLLARD, J. Hostility and fear in social life. In T. M. Newcomb & E. L. Hartley (Eds.), *Readings in social psychology*. New York: Holt, 1947. Pp. 269–281.

EDWARDS, A. *Experimental design in psychological research*. New York: Rinehart, 1950.

ELIZUR, A. Content analysis of the Rorschach with regard to anxiety and hostility. *J. proj. Tech.*, 1949, **13**, 247–284.

GIBBY, R. The stability of certain Rorschach variables under conditions of experimentally induced sets: I. The intellectual variables. *J. proj. Tech.*, 1951, **15**, 3–26.

GORLOW, L., ZIMET, C., & FINE, H. The validity of anxiety and hostility Rorschach content scores among adolescents. *J. consult. Psychol.*, 1952, **16**, 73–75.

JOEL, W. The interpersonal equation in projection methods. *Rorschach Res. Exch.*, 1949, **13**, 479–482.

KIMBLE, G. A. Social influence on Rorschach records. *J. abnorm. soc. Psychol.*, 1945, **40**, 89–93.

KLATSKIN, ETHELYN. Effect of the test situation upon the Rorschach record. *J. proj. Tech.*, 1952, **16**, 193–199.

LORD, EDITH. Experimentally induced variations in Rorschach performance. *Psychol. Monogr.*, 1950, **64**, No. 10 (Whole No. 316).

MCNEMAR, Q. *Psychological statistics*. New York: Wiley, 1949.

MILLER, D. R. Prediction of behavior by means of the Rorschach test. *J. abnorm. soc. Psychol.*, 1953, **48**, 367–375.

SANDERS, R., & CLEVELAND, S. E. The relationship between certain examiner personality variables and subjects' Rorschach scores. *J. proj. Tech.*, 1953, **17**, 34–50.

SARASON, S. B. The test situation and the problem of prediction. *J. clin. Psychol.*, 1950, **4**, 387–392.

SCHACHTEL, E. G. Subjective definitions of the Rorschach test situation and their effect on test performance. Contributions to an understanding of Rorschach's test. *Psychiatry*, 1945, **8**, 419–448.

SINGER, J. L., MELTZOFF, J., & GOLDMAN, G. D. Rorschach movement responses following motor inhibition and hyperactivity. *J. consult. Psychol.*, 1952, **16**, 359–364.

TOWBIN, A. P. Hostility in Rorschach content and overt aggressive behavior. Unpublished doctoral dissertation, Yale Univer., 1955.

WALKER, R. G. A comparison of clinical manifestations of hostility with Rorschach and MAPS test performance. *J. proj. Tech.*, 1951, **15**, 444–460.

Received January 14, 1958.

THE GENERALITY OF SOCIAL PERCEPTION SCORES[1]

MOHAMMED Y. QUERESHI
University of Illinois

RECENT studies of interpersonal perception (e.g., Bender & Hastorf, 1950; Bieri, 1953; Quereshi, 1957) claim to investigate such constructs as transference, empathy, real similarity, insight, etc., in terms of their operational equivalents. They have typically employed distance scores such as D^2 or D (Cronbach & Glaser, 1953) or simple absolute difference on the assumption that (*a*) the so-called operational measures of social perception are general over dimensions regardless of the content, and that (*b*) the sign of the difference score, i.e., the direction of the difference, is not important. A study bearing on the first of these assumptions was carried out by Cronbach, Hartman, and Ehart (1953), who report that "assumed similarity," indexed by absolute difference between two "cluster" scores, is a general attitude or mental set, independent of item content. Since they categorized items into clusters on the basis of logical analysis, however, their procedure could not insure independence of contents, and their claim that the "assumed similarity" measure was general over content-fields cannot be accepted as conclusive. In addition, few attempts have been made to estimate the reliability (internal consistency) of these scores with due regard to the content.

The present study was designed to break ground in this area by first analyzing the perceptual data into a number of orthogonal dimensions and then estimating (*a*) the generality and (*b*) the internal consistency reliability of a variety of measures of the perceptual relations (to be described in detail later). For each measure there were three main questions to be answered:

1. Is the measure reliable (consistent) within dimensions, i.e., consistent over items representing a particular content-field? If so, then to what extent?

2. Is the measure based on orthogonal factors general over various dimensions or content-fields? If so, then to what extent?

3. What is the effect of disregarding the sign of a difference or partial score on its generality over dimensions and its consistency within dimensions?

Answers to these questions have both methodological and practical significance for research in social perception. They should also throw light on conflicting results and confusion that exist in the present literature of this field.

METHOD

Subjects

During the initial stages of a program of research on therapeutic counseling at the Student Counseling Bureau of the University of Illinois, a number of tests, rating scales, and questionnaires were administered to persons who sought help from the Counseling Bureau. One of the rating scales administered to 204 students, 147 males and 57 females, was Ewing's Personal Rating Form, briefly described below. This sample consisted mostly of those who came to the Counseling Bureau for some sort of psychological help for their personal and emotional problems. However, some of the *S*s came from an undergraduate psychology course, and we can assume that they were relatively well-adjusted.

Instruments and Data

The Ewing Personal Rating Form[2] consists of 100 adjectives such as "steady," "frank," "repressed," etc. (Quereshi, 1957). Each *S* was provided with three copies of this check list and was asked to rate himself, his father, and his mother on a five-point scale ranging from "Not Typical" to "Typical." The first part of the study was based on the self-ratings of the entire reference group of 204 *S*s. Other-ratings (father and mother) were available for 98 males out of the total 204. The

[1] This paper is based on the author's master's thesis submitted to the Graduate College of the University of Illinois in 1957. Deep appreciation is due to Lee J. Cronbach for suggesting the problem and providing helpful suggestions and guidance throughout the course of this study. Thanks are also due to William Hays of the University of Michigan for a critical reading of the manuscript.

[2] A 9-page document containing the table of rotated factor loadings, the Ewing Personal Rating Form, and keys for all the factors, has been deposited with the American Documentation Institute. Order Document No. 5809 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D.C., remitting $1.25 for microfilm or $1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

self-ratings and father-ratings of this smaller group formed the basis for the second part of the study.

Let the self-rating score on an item i for an individual a be designated as X_{ia}, and the father-rating score on the same item for the same individual as Y_{ia}. A variety of scores used in this study are based on or derived from these rating scores:

Factor score. The perceptual contents were to be analyzed first into orthogonal factors or dimensions. Each orthogonal factor was to be represented by a group of items. Thus a factor score for an individual a is equal to the sum of his rating scores on all the items defining a factor: These scores may be represented as

$$\sum_{i=1}^{k} X_{ia} \quad \text{or} \quad \sum_{i=1}^{k} Y_{ia}$$

depending upon whether we are using self-ratings or father-ratings to obtain the factor score. The respective factor scores may be designated C_{fsa} and C_{fya}.

Factor difference score with regard to sign. This score is obtained by subtracting the factor score based on self-ratings from the factor score based on father-ratings, i.e.,

$$d_{sa} = \left(\sum_{i=1}^{k} Y_{ia} - \sum_{i=1}^{k} X_{ia} \right)$$

Factor difference score without regard to sign. This is obtained the same way as the immediately preceding score above, except that here the sign of the difference is disregarded. This score can be represented symbolically as

$$| d_{sa} |$$

Factor partial score with regard to sign. A factor partial score for an individual is based on the difference between the actual factor score based on father-ratings of that individual and the father-ratings factor score as predicted from the self-ratings factor score. Thus a factor partial score may be represented by:

$$P_{sa} = C_{fya} - \left\{ \bar{C}_{fya} + r_{C_{fya}C_{fsa}} \frac{\sigma C_{fya}}{\sigma C_{fsa}} (C_{fsa} - \bar{C}_{fsa}) \right\}$$

Factor partial score without regard to sign. This is obtained in the same way as the immediately preceding score, except that in this case the sign of the partial score is disregarded, i.e.,

$$| P_{sa} |$$

Factor D^2 and D score. The factor D^2 for an individual is obtained by squaring the difference between father-ratings and self-ratings for that individual and summing them over items comprising a factor. Thus, the D score for an individual a for factor f is

$$D_{fa} = \sqrt{\sum_{i=1}^{k} (X_{ia} - Y_{ia})^2}$$

Procedure

The first part of this study consisted of factor analysis of covariances based on self-ratings of the 204 *Ss*. Each item was treated as a variable, and covariances were computed for the self-ratings. The matrix of covariances was then analyzed, so as to extract orthogonal dimensions by means of Thurstone's Centroid Method. As many of the factors as possible were retained in order to account for the maximum possible covariance. Out of 10 unrotated factors, 7 were retained and subjected to quartimax rotation in order to approximate simple structure. Examination showed that quartimax rotation did not approach the desired degree of psychological meaningfulness in the case of Factors V and VII, which were therefore rotated graphically to improve meaningfulness.

Inspection of the item factor loadings after rotation showed that a number of them were extremely low. Some of the items could thus be excluded from the later operations without incurring the loss of any substantial covariance. Factor keys were therefore composed from positive or negative loadings, i.e., loadings $\pm.50$ or higher on Factor I and $\pm.30$ or higher on Factors II through VII. These keys formed the basis for the factor scores on self- and father-ratings for the 98 male *Ss*. The items thus included in the factor keys[3] gave unit positive or negative weight to each raw score, since an empirical comparison by Trites and Sells (1955) between this unit weighting method and fractional weighting concluded that little could be gained by resorting to the latter.

In order to obtain estimates of the internal consistency, each factor key was subdivided by sorting the selected items into two equated groups, and assigning items to half-factor keys (A and B) in such a manner that they were fairly balanced with respect to number of items and magnitude and sign of loading.

Both the full-factor scores and the half-factor scores (based on items in either key A or B of a factor) were then computed for all the factors, and the pairwise correlations among the 14 half-factor scores determined, both for the self-ratings and for the father-ratings. Half- and full-factor difference and partial scores were next computed, both with and without regard to sign.

The Item D scores were also computed for both the half-factor keys and the full-factor keys. This procedure yielded 14 half-factor Item D scores and 7 full-factor Item D scores.

Correlations were computed among half-factor as well as full-factor scores of all the types mentioned above. These provided evidence both on the reliability (internal consistency) of these scores for various content-fields or dimensions and their generality across dimensions.

RESULTS AND DISCUSSION

As already reported, the item covariances were factor analyzed using Thurstone's centroid method and quartimax rotation. Seven

[3] The number of items in each factor key is as follows: Factor I = 24; Factor II = 24; Factor III = 25; Factor IV = 23; Factor V = 17; Factor VI = 17; Factor VII = 11.

TABLE 1
SPLIT-HALF RELIABILITIES, MEANS AND STANDARD DEVIATIONS OF VARIOUS SCORES USED IN THIS STUDY ($N = 98$)

Scores	Factor I (Unhappy)				Factor II (Outgoing)				Factor III (Contrary)				Factor IV (Hardworker)			
	r	r^*	M	SD	r	r^*	M	SD	r	r^*	M	SD	r	r^*	M	SD
Scores based on self-ratings	.89	.94	63.5	15.4	.65	.79	31.6	12.2	.84	.91	42.6	13.6	.75	.86	52.7	11.4
Scores based on father-ratings	.87	.93	45.8	16.3	.56	.72	31.7	11.3	.85	.92	47.2	17.3	.85	.92	56.2	14.7
Difference scores with regard to sign	.81	.90	−17.7	17.3	.55	.71	00.0	15.3	.80	.89	4.6	19.8	.83	.91	3.5	17.7
Absolute difference scores	.74	—	20.2	14.3	.38	—	11.9	9.7	.57	—	15.9	12.7	.70	—	14.1	11.2
Partial scores with regard to sign	.84	—	00.0	14.9	.55	—	00.0	11.3	.83	—	00.0	17.0	.86	—	00.0	14.6
Absolute partial scores	.64	—	12.1	8.7	.37	—	8.7	7.1	.69	—	13.8	9.9	.74	—	11.3	9.2
D scores	.74	—	6.7	4.2	.59	—	6.2	3.5	.63	—	6.6	4.1	.78	—	5.7	3.9

Scores	Factor V (Conventional)				Factor VI (Cyclothyme)				Factor VII (Creative)			
	r	r^*	M	SD	r	r^*	M	SD	r	r^*	M	SD
Scores based on self-ratings	.45	.62	22.2	7.1	.70	.82	−15.9	8.9	.47	.64	22.7	5.3
Scores based on father-ratings	.42	.59	22.5	7.4	.75	.86	−14.2	10.8	.57	.73	19.8	5.4
Difference scores with regard to sign	.49	.66	00.3	10.8	.73	.84	1.8	12.5	.45	.62	−2.9	6.0
Absolute difference scores	.03	—	8.4	6.8	.49	—	10.0	7.7	.24	—	5.1	4.3
Partial scores with regard to sign	.40	—	00.0	7.3	.76	—	00.0	10.6	.53	—	00.0	5.0
Absolute partial scores	.02	—	6.0	4.3	.54	—	8.5	6.3	.09	—	4.2	2.6
D scores	.49	—	4.2	2.7	.55	—	4.0	2.4	.30	—	2.4	1.5

r^* Corrected by Spearman-Brown Formula.

factors that were selected for later investigation have been named as follows:

Factor I	Unhappy
Factor II	Outgoing
Factor III	Contrary
Factor IV	Hard worker
Factor V	Conventional
Factor VI	Cyclothyme
Factor VII	Creative

It was apparent from the beginning that the factor scores thus obtained would not be orthogonal to one another, for one reason because some items were included in more than two factor keys. Such a case is illustrated by Factors III and VI. Since these two factors are highly correlated with one another, Factor VI may either be dropped or combined with Factor III in subsequent research without losing any significant information. On the whole, as the figures for Scores 1 and 2 indicate, the seven factors have a median absolute intercorrelation of approximately .30. Orthogonality of factor scores could only have been achieved in this study by the laborious assignment of precise item weights.

Reliability (Internal Consistency) of Various Scores

Table 1 presents the split-half reliabilities, means, and standard deviations of the various scores. For the first three types of scores, reliability has been corrected by the Spearman-Brown formula and is reported together with the uncorrected estimates. For the last four scores, the use of the Spearman-Brown formula is questionable because these scores do not meet its assumptions.[4]

The reliability estimates (Table 1) for factor scores based on self-ratings and for factor scores based on father-ratings are approxi-

[4] For example, in the case of D score, the correlation between two parallel keys of double length would be

$$r\sqrt{D^2_{fa} + D^2_{fa}}\sqrt{D^2_{fa1} + D^2_{fa1}}$$

whereas the S-B equation estimates

$$r_{(D_{fa}+D_{fa})(D_{fa1}+D_{fa1})}.$$

TABLE 2

MEDIAN AND MEAN ABSOLUTE INTERCORRELATIONS OF VARIOUS SCORES
(Estimates of factor-intercorrelations and generality of scores)

Scores	Factor I (Unhappy) with the other 6 factors		Factor II (Outgoing) with the other 6 factors		Factor III (Contrary) with the other 6 factors		Factor IV (Hard worker) with the other 6 factors	
	Median	Mean	Median	Mean	Median	Mean	Median	Mean
Factor scores based on self-ratings	.35	.27	.36	.32	.21	.27	.13	.19
Factor scores based on father-ratings	.36	.37	.32	.35	.34	.42	.37	.36
Factor difference scores with regard to sign	.31	.25	.36	.36	.26	.36	.28	.29
Factor absolute difference scores	.09	.12	.19	.21	.18	.25	.11	.12
Factor partial scores with regard to sign	.37	.33	.34	.36	.35	.42	.39	.37
Factor absolute partial scores	.13	.19	.20	.20	.24	.29	.11	.15
Factor *D* scores	.56	.56	.72	.72	.52	.63	.66	.64

Scores	Factor V (Conventional) with the other 6 factors		Factor VI (Cyclothyme) with the other 6 factors		Factor VII (Creative) with the other 6 factors	
	Median	Mean	Median	Mean	Median	Mean
Factor scores based on self-ratings	.27	.29	.25	.31	.09	.17
Factor scores based on father-ratings	.34	.34	.35	.45	.27	.29
Factor difference scores with regard to sign	.35	.31	.29	.39	.24	.23
Factor absolute difference scores	.34	.29	.14	.25	.20	.19
Factor partial scores with regard to sign	.33	.33	.38	.44	.23	.28
Factor absolute partial scores	.25	.22	.17	.27	.08	.12
Factor *D* scores	.70	.68	.63	.63	.63	.62

mately the same.[5] Also, the reliability estimates for factor difference and partial scores with regard to sign are strikingly close to one another and almost identical with those for factor scores based on self- and father-ratings, respectively. The corrected reliability estimates for the first three types of scores listed in the table, for four out of seven factors, are in the vicinity of .90. For the remaining three factors (II, V, and VII) they range between .59 and .79.

It is significant that the uncorrected reliability estimates for absolute difference and partial scores for all factors are consistently lower than the uncorrected reliability estimates for the same scores when the sign is considered. In every case, the reliability estimates for the former scores are lower than those for the latter, and in some cases, e.g., Factors V and VII, the reductions in magnitudes are significant at .05 or less. Ignoring the sign of the difference or partial score thus results in loss of information.

The split-half reliability estimates for *D* scores are generally higher than those for the absolute difference and partial scores, but are consistently lower than those for the other four types of measures for all factors except II and V. For these two factors, the reliabilities are either close to those for these four measures or a little higher. On the whole, the tendency is toward a decrease in the magnitude of reliability for *D* scores when compared with the four types of scores referred to above, though further investigation is needed on this point.

Generality of Various Scores

Since the basic factor scores are intercorrelated, the derived scores based on them should also be intercorrelated by statistical necessity. The results in Table 2 show that the

[5] It may be pointed out that the reliability of the difference scores with regard to sign could be computed from the reliability of full-factor scores based on self-ratings and of those based on father-ratings by the simple use of formulas available for this purpose. The values thus obtained would agree with those we have calculated directly.

TABLE 3
INTERCORRELATIONS OF SCORES FOR PAIRS OF ORTHOGONAL FACTORS

Factor Pair	Intercorrelations of Scores*						
	s	*f*	*d*	*/d/*	*P*	*/P/*	*D*
I with VII	−.08	−.17	−.03	.10	−.12	.10	.61
III with VII	−.02	−.18	−.19	.20	.01	.09	.52
I with V	.08	−.02	−.28	.09	−.21	.01	.49
IV with VI	.11	−.30	−.22	−.07	−.31	−.00	.55
IV with V	−.06	−.21	−.09	.12	−.20	.24	.72
III with IV	.14	−.16	−.15	.02	−.18	.11	.52
II with VI	.17	.09	.23	.08	.13	−.08	.71
VI with VII	−.09	−.27	−.13	.20	−.21	.05	.54

* *s* stands for correlation based on self-ratings; *f* for correlation based on father-ratings; *d* for difference score with regard to sign; */d/* for difference score without regard to sign; *P* for partial score with regard to sign; */P/* for partial score without regard to sign; and *D* for D score as explained elsewhere.

Note.—The above correlations are based on 98 cases, and p for an r of .20 = .05.

median absolute correlations representing the generality of factor difference and of factor partial scores (both with and without regard to sign) are hardly different from those of the factor scores based on self-ratings and father-ratings. Since the last two scores indicate the degree to which the factors are intercorrelated, the factor difference and factor partial scores would appear to be general only to the extent to which the perceptual contents overlap.

The *D* scores, unlike the factor difference and partial scores with regard to sign, show appreciable generality over dimensions. Even after making due allowance for the intercorrelations of factors on which *D* scores are based, the *D* scores are significantly general over dimensions. The difference between the median of median *D* score absolute interfactor correlations (.63) and the median of median absolute interfactor correlations based on factor scores for self- and father-ratings (.33) is statistically significant at the .05 level. In the light of the available evidence, the *D* score can thus be justifiably used in research on social perception without regard to the content dimensions involved.

One of the initial questions concerned the effect of ignoring the sign of a difference or partial score on its generality. A comparison of the median *r*'s for factor difference and partial scores with regard to the sign with the median *r*'s of the absolute difference and partial scores indicates a consistent and appreciable drop in generality for all the factors. In the case of partial scores, for some factors the drop in generality estimates (from .37 to .13 for Factor I, and from .39 to .11 for Factor IV, respectively) is significant at .05 level (*P* of a difference of .24 between two correlated *r*'s being $\leq$ .05). These results definitely lend support to the view that the sign of a difference or partial score is important for its generality and that ignoring it results in the loss of relevant information.

The drop in generality estimates of difference and partial scores, when the sign is disregarded, is also apparent in Table 3, which shows the intercorrelations of scores for pairs of orthogonal factor scores. Here the correlations are reported with their sign, unlike Table 2 which showed medians of absolute intercorrelations of factors. The estimates of correlations for most pairs are appreciably reduced when the sign of the score is disregarded. This tendency, however, is not as sharply illustrated in these results as in those of Table 2. The *D* score for these pairs of factors is significantly general, the *r*'s ranging between .52 and .72.

An Evaluation

Since the method of analyzing social perception scores adopted in this study has been tried here for the first time, it would be premature to assert that its usefulness for research in this field has been demonstrated. Moreover, the approximation of factor scores by factor keys employing unit weights is also a preliminary technique subject to modification and refinement. Caution is in order in interpreting the results concerning the reliability and generality of various scores used in research on interpersonal perception, and concerning the effect of ignoring the direction of difference on their reliability and generality. However, it

can be justifiably stated that the evidence so far available does not support the generality of most of these scores, and indicates that the sign of difference or partial measures can be ignored only with the loss of relevant information.

The generality of *D* scores, as supported by the present investigation, has important bearing on their use in social perception research. These results lend support to the view that the *D* score can be used without resorting to dimensional analysis.

SUMMARY

This study of self-ratings and other-ratings was designed to investigate (*a*) whether it is justified to assume the generality of various distance scores regardless of the content involved, (*b*) the reliability of these scores, and (*c*) the effect on their generality and reliability when the sign of any of the distance measures is ignored. The method of data analysis employed is comparatively new and has been tried here for the first time.

The results indicate that what generality most of these scores possess can be attributed to overlap between the content areas on which the ratings are based. Except for one of the measures (*D* score) used in this study, all follow the same pattern. Appreciable drop in generality is consistently found when the sign of difference is disregarded. The generality of *D* score seems to justify its use as a defensible global measure.

Reliability estimates are also appreciably reduced in magnitude when the signs of the distance measures are not taken into consideration (the *D* score is no exception here). The results do not approach desired statistical significance in all cases, but the trends are clearly and consistently indicative of the loss of relevant information when the sign of the measure is disregarded.

REFERENCES

BENDER, J. E., & HASTORF, A. H. The perception of persons forecasting another person's responses on three personality scales. *J. abnorm. soc. Psychol.*, 1950, **45**, 556–561.

BIERI, J. Changes in interpersonal perception following social interactions. *J. abnorm. soc. Psychol.*, 1953, **48**, 61–66.

CRONBACH, L. J. Proposals leading to analytic treatment of social perception scores. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior*. Stanford, Calif.: Stanford Univer. Press, 1958. Pp. 353–379.

CRONBACH, L. J. Processes affecting scores on "understanding of others" and "assumed similarity." *Psychol. Bull.*, 1955, **52**, 177–193.

CRONBACH, L. J., & GLASER, GOLDINE C. Assessing similarity between profiles. *Psychol. Bull.*, 1953, **50**, 456–474.

CRONBACH, L. J., HARTMAN, W., & EHART, MARY E. Investigation of the character and properties of assumed similarity measures. Bur. of Educ. Res., Univer. of Illinois, 1953. (Mimeo.)

QUERESHI, M. Y. An empirical study of the generality of social perception scores. Unpublished master's thesis, Univer. of Illinois, 1957.

TRITES, D. K., & SELLS, S. B. A note on alternative methods for estimating factor scores. *J. appl. Psychol.*, 1955, **39**, 455–456.

Received December 30, 1957.

SUBJECTIVE CERTAINTY AND RESISTANCE TO CHANGE[1]

DONALD S. BOOMER

National Institute of Mental Health

THIS study is concerned with the psychological bases for resisting or conforming to group influence. The point of departure was the desire to tap subjects' "private" experience during the application of group pressure in order to learn more about the psychological processes involved.

Experimental approaches to conformity most typically identify conforming subjects by some simple behavioral criterion, e.g., revising a discrepant estimate or changing a vote to agree with a group consensus. Dichotomizing subjects (*S*s) thus according to their behavior at a choice point lends itself readily to the assumption that the underlying psychological processes are similarly dichotomized. Such an assumption ignores the fact that the overt behavior being noted is the only observable aspect of a complex psychological process, and identical behavior among subjects does not necessarily reflect a corresponding identity, or even similarity, in the psychological processes involved.

This principle has been cogently demonstrated by Asch (1952) in investigating the problem of the "minority of one vs. a unanimous majority." His basic experimental design categorized *S*s behaviorally as "independents" and "yielders." Postexperimental interviews, however, revealed among his *S*s three qualitatively different kinds of processes that led to yielding and two that led to maintaining independence. These differences had not been discernible in the behavioral phase of the experiment.

Our own study employed periodic estimates of certainty[2] by the *S* during the period of experimental group pressure in order to see what differences they might reveal among people whose observable behavior was similar. The procedure was adapted from Crutchfield (1951) who in turn had adapted it from an unpublished problem-solving task devised by Bavelas. A brief summary of Crutchfield's procedure provides the necessary context for a statement of the central issue of the present study.

His *S*s worked in groups of six on a group problem that involved fitting together, jigsaw-fashion, irregularly shaped cardboard pieces into squares. Each *S* had one such square to make; the task required each *S* to complete his square from the limited set of cardboard pieces available in common to the group. The *S*s were told that they would be exchanging pieces with one another in order to complete their squares.

This process of exchange was accomplished indirectly through the experimenter, enabling him to maintain strict control of the situation. Actually no real transactions were permitted; the experimenter posed requests and made pieces available for exchange according to a preset schedule, maintaining the illusion for the *S*s that they were dealing with one another.

Each *S* was permitted to complete his square early in the experiment, but was led to believe that others were not yet finished, since the experimenter continued to make rounds of the group presenting simulated "requests" for pieces still needed. After a few such circuits of the group, the experimenter presented a request for one of the pieces that each *S* was using in his completed square. The crucial determination was how many successive requests were required of a *S* before he broke his square to yield the piece requested.

Some *S*s "broke" at the first such request; at the other extreme one *S* broke only after 20 successive requests. Crutchfield (1951) grouped his *S*s into three categories: fast, medium, and slow breakers. In his report he says, "The test confronts each person in the group with an identical problem such that he is placed in a conflict arising from competing demands of cooperation, of independent analysis of the problem, and of maintenance of a

[1] This research is a portion of a Ph.D. thesis submitted to the Department of Psychology, University of California. The author is indebted to the members of his doctoral committee, Read D. Tuddenham, Richard S. Crutchfield, and Gordon W. Hearn, for their guidance and criticism.

[2] The writer is indebted to Jack Block for suggesting this specific measure.

stable configuration. The person's resulting behavior in the conflict situation is objectively and simply scored" (pp. 587 f.). The inferential reconstruction of the psychological situation is then temptingly straightforward. The first request for the crucial piece generates this conflict; each *S* is torn between the desire to cooperate and the desire to maintain his stable configuration. Whichever of these motives is the more potent determines whether he will break early or late.[3]

Consideration of Asch's interview findings and James' Psychologists' Fallacy raises questions about this interpretation. Part of the variability among these *S*s might be due to their relative readiness to become aware that the situation posed a conflict for them. Some slow breakers might be slow not because the stability of their own solutions was more important to them than cooperation, but because they had protected themselves from experiencing this conflict by their perceptual organization of the situation. Data were sought concerning the *S*s' periodic assessments of their own feeling of confidence during the experiment in the hope that they would indicate when and if *S*s were experiencing conflict. The larger study (Boomer, 1954) on which this report is based was also concerned with some other parameters of Crutchfield's experimental situation which will not be reported here.

METHOD

Subjects

The *S*s were 72 undergraduate college students, 36 men and 36 women, drawn from a large lower division course in personal and social adjustment. They were not volunteers; all *S*s were required to participate in one experiment as a part of their course work. These *S*s were randomly divided into 12 experimental groups: six groups of six men and six groups of six women.

Procedure

The experimental procedure was modelled quite closely on the Crutchfield (1951) "broken square" procedure described above, with the modification that numbers were substituted for jigsaw pieces as the experimental materials. Relevant parts of the instructions, adapted directly from Crutchfield's report, are reproduced here in order to explain the procedure and to give the flavor of the situation. These instructions were read aloud to the experimental groups of six *S*s each:

> This is a group problem-solving situation with a group goal. There are no individual goals in it separate from the group goal. The six members of this group will work on a common group problem, but the only communication among you will be via the experimenter, who will circulate between you throughout the task, serving as a messenger.
>
> The materials are a set of small numbered wooden discs. Each of you will begin with five randomly selected numbers, and the group task is for each of you to acquire by trading back and forth either a sequence of five numbers—5–6–7–8–9, say, or a set of five numbers, five 3's, say. Each person's part of the total group problem is to assemble one such set or sequence. The group goal is considered achieved only at that point when *all six* of you have a set or a sequence.
>
> The available numbers range from 1 to 15, but you have no way of knowing how many of each there are. Actually the numbers have been so chosen that there are many ways for one or more of you to acquire a set or a sequence, but very few ways in which all of you can get simultaneous solutions. This is the real essence of the group task.
>
> To make this point clear I will show you a similar set-up in which the task is to acquire a set or a sequence of only three numbers. (The following diagram is drawn on the blackboard:)
>
> | | |
> |---|---|
> | A. 1 2 3 | D. 6 6 6 |
> | B. 2 3 4 | E. 7 7 7 |
> | C. 3 4 5 | F. 9 9 9 |
>
> These six people, represented here by the letters A, B, C, etc., have traded their original numbers back and forth until each of them has a solution. This is one of the few ways in which this set of numbers could be arranged to provide six simultaneous solutions. If someone had chosen to collect 3's, however, it would have been impossible for A and B to get sequences. Or had someone chosen to save this sequence, 5–6–7, it would have been impossible for D and E to get sets.

All exchanging of numbers took place via the experimenter (*E*), who carried a small tray from *S* to *S*. One section of the tray held numbers that were available for exchange: another section of the tray showed numbers that had been purportedly requested by members of the group. This provision for indirect communication through *E* permitted the deception on which the experiment depended.

*S*s were seated in a semicircle, facing outward

[3] In the interests of clarity, the third factor, "independent analysis of the problem," is not considered, since the data on estimation of certainty cannot be brought to bear upon this factor as a possible aspect of the conflict.

so they could not see one another directly nor *E*, who walked behind them circulating the tray. *E* was thus free to manipulate the numbers on the tray in such a way that a standard stimulus situation could be maintained for all *S*s throughout the experiment. On each round, or circuit of the group, *S*s all faced an identical exchange-request situation. When an *S* exchanged a piece or requested one, *E* secretly restored the standard situation before presenting the tray to the next *S*. Thus, although each *S* believed he was freely communicating with the other members of the group he was actually dealing only with *E* in a controlled situation.

The sequence of events was this: All *S*s began the task with identical "hands," 4–5–7–7–10, which they were led to believe was a chance draw. By offering a 6 at the first round and an 8 at the third, *E* led each *S* to build the sequence, 4–5–6–7–8. Rounds 4 and 5 offered and requested numbers above 10, which did not concern any *S*.[4]

At Round 6 each *S* was confronted with a request, ostensibly from a teammate, for a 7. Those *S*s who broke their sequence to yield the 7 were permitted to build a new sequence, 2–3–4–5–6, during the next two rounds, while the pressure of the critical request was maintained on each round for the others. In the event that some *S*s in the group still had not broken their sequences after 15 crucial rounds, the experiment was discontinued, and the group was told that they had succeeded in the group task. This was credible since each *S* had a solution, either his original 4–8 sequence or the replacement 2–6 sequence.

Each *S* was also furnished at the outset with a pad of small protocol slips numbered to correspond with the rounds. He filled out the appropriately numbered slip after every round indicating what trade he had made, if any, and what number he had requested, if any. If he had neither traded nor requested, he checked "pass" on the slip. On each round of the group, *E* collected the slips from the previous round, checking them for accuracy. These provided a succinct record of each *S*'s transactions on each round.

In addition to recording his transactions, *S* estimated his feeling of confidence[5] at each round. At the bottom of each protocol slip was a three-inch line with 0 at the left, 50 in the center and 100 at the right. Along this line *S* was asked to write a number representing his degree of certainty that his transaction on this round had been "the best thing to do" for him and his group. These certainty estimates will hereinafter be referred to as CEs.

A final source of information was the free comments which *S*s were urged to write on the backs of these slips. Since *E* got around the group only once every 3 or 4 minutes, *S*s had ample time to formulate their impressions. These written comments were not analyzed as data, but did yield some useful impressions about *S*s' perception of the unfolding situation.

RESULTS

The number of the trial at which *S* yielded to the request for a 7 will be termed his "break score." Break scores ranged from 6, the first request round, to 20, the last round.[6] Twenty-three *S*s had not yielded at the last round; their scores were recorded as NB (nonbreak).

The distribution thus obtained, like Crutchfield's, was markedly skewed. The following classification of scores was employed in the statistical analysis:

1. "Early breakers," $N = 17$, *S*s yielding on Rounds 6 and 7, the first two request rounds
2. "Moderate breakers," $N = 17$, *S*s yielding on rounds 8 and 9
3. "Late breakers," $N = 13$, *S*s yielding on rounds 10 through 20
4. "Nonbreakers," $N = 23$.

Certainty estimates. The mean CEs of nonbreakers and breakers are generally discrepant throughout the experiment, most markedly at certain critical junctures. Table 1 shows the mean CEs for these two groups over the first six rounds. Beginning at round three the mean

[4] These two noncrucial rounds were deliberately inserted to provide for the possibility that some *S*s might not accept the forced sequence on Rounds 2 and 3. Only three *S*s failed to complete the sequence at Round 3, and they completed it on Round 4.

[5] Although these determinations were originally termed "confidence estimates," they are called "certainty estimates" in this report to avoid possible confusion with the statistical concept.

[6] Two *S*s unaccountably broke their sequences on Round 5, just before the crucial rounds began. Since their subsequent experience was not comparable to that of the other *S*s, their data were discarded, reducing the N to 70.

TABLE 1

MEAN CERTAINTY ESTIMATES OF BREAKERS AND NONBREAKERS FOR ROUNDS 1 THROUGH 6

Round	Breakers		Nonbreakers		t	p
	N[a]	$\bar{X}$	N[a]	$\bar{X}$		
1	47	59.0	22	70.8	1.76	NS
2	47	47.8	22	60.4	1.96	NS
3	46	73.6	21	96.8	5.57	.001
4	45	77.7	21	92.2	2.96	.01
5	42	78.4	21	91.4	2.65	.02
6	36[b]	72.0	22	86.6	2.45	.02

[a] The discrepancies in the Ns for Rounds 1–5 are due to Ss' occasional failure to record CEs on their protocols.

[b] CEs of the 10 Ss who broke on this round are not included in this mean.

TABLE 2

FREQUENCY OF SUBJECTS IN FOUR SCORE CATEGORIES EXPRESSING CERTAINTY ESTIMATES OF 100 AT ROUND 3

Score Category	Number of Subjects Expressing Certainty Estimates of 100	Number of Subjects Expressing Certainty Estimates of < 100	Totals
Early breakers	4	12	16
Moderate breakers	5	12	17
Late breakers	6	7	13
Nonbreakers	19	3	22
Totals	34	34	68[a]

Note.—Chi square = 19.28; $p < .001$.

[a] Two Ss neglected to record any certainty estimate on this round.

CEs for breakers and nonbreakers are significantly different.

Contributing heavily to their higher mean CEs was the nonbreakers' more frequent use of 100 as an estimate. This finding suggested another approach to these data. The difference between 100 and any smaller number, even 99, represents a distinction between the expression of absolute certainty and the admission of some uncertainty. Let us consider the relative incidence of CEs of 100 in the various break-score categories.

Two junctures are presented as being particularly salient in reflecting differential response to the experimental situation. The first is Round 3, when S completed his first sequence. Table 2 presents the frequency of 100% CEs on this round for each of the four break-score categories. The frequency differences among the categories yield a chi square which is significant at the .001 level of confidence. It is clear that whether or not S expressed complete confidence at this very early point is closely related to whether or not he will ultimately yield.

TABLE 3

FREQUENCY OF SUBJECTS IN FOUR SCORE CATEGORIES EXPRESSING CERTAINTY ESTIMATES OF 100 AT ROUND 6

Score Category	Number of Subjects Expressing Certainty Estimates of 100	Number of Subjects Expressing Certainty Estimates of < 100	Totals
Early breakers	0	6	6[a]
Moderate breakers	5	12	17
Late breakers	5	8	13
Nonbreakers	13	9	22
Totals	23	35	58[b]

Note.—Chi square = 8.17; $p < .05$.

[a] CEs of the 10 Ss who broke on this round are not included in this table.

[b] Two Ss neglected to record any certainty estimate on this round.

The other juncture selected for scrutiny was Round 6, when S first encountered the request for the 7 he was using in his sequence. The CE on this round reflects S's initial reaction to this, the crucial conflict posed by the experimental situation. Ten Ss yielded to this first request; their CEs are excluded from consideration here, since their actual situation was changed by their having broken. The remaining 35 breakers behaved like the nonbreakers on this round in that they maintained their solutions despite this initial request, but they expressed less certainty about it than did the nonbreakers. The relative frequencies of CEs of 100 for the four groups are shown in Table 3. The frequency difference among these groups yields a chi square which is significant at the .05 level of confidence.

DISCUSSION

To summarize the CE data, they demonstrate that Ss who express perfect certainty at Round 3, upon completing the initial sequence, are statistically less likely to break this sequence later than are those who express some uncertainty at this juncture. Similarly, at Round 6 those Ss expressing 100% certainty when declining to yield to this first request for the 7 are less likely to yield on any successive

round than are those who express some uncertainty.

If all *S*s were indeed experiencing the presumed conflict between the demands of cooperation and of maintaining a stable configuration there would be no reason to expect systematic CE differences between breakers and nonbreakers. If the demand for the 7 at Round 6 in fact poses a dilemma for all *S*s, this should be reflected in lowered CEs among nonbreakers as well as breakers. Instead the data show that some *S*s express no uncertainty about maintaining their solutions intact in the face of this request and that these *S*s are predominantly nonbreakers.

The instructions, it will be recalled, do not explicitly anticipate this situation in which *S* is asked for a number already imbedded in his completed sequence. When *S* is confronted with this situation at Round 6, its impact on him depends on his perceptual organization of the total situation, including his recall of the instructions and the meaning he places on his own completed sequence.

Differences among *S*s with regard to the meaning of having achieved a sequence are reflected in their CEs at Round 3. The expression of perfect certainty at this point is unwarranted, considering *all* the information which was provided in the instructions. It had been clearly stated that the task was a group task, that many individual solutions were possible but very few simultaneous group solutions. It was even explicitly demonstrated how one premature individual solution could block the group solution.

If *S* fails to recall this critical aspect of the task he can view his personal solution as the completion of his part of the group task and register 100% certainty that completing this sequence was the best thing to do for himself and his group. With such a set he can, without discomfort, interpret the subsequent request for a 7 as not directed to him, and he can keep his solution intact, still with 100% certainty that this is the best course. Such *S*s are not in conflict between the competing demands of cooperation and of maintaining a stable configuration; they have perceptually organized the situation in such a way as to eliminate conflict.

Support for this interpretation of the psychological implications of high and low CEs was provided by the free written comments of some of the more verbal *S*s. Following are clear statements by four *S*s at Round 3, when each had completed his sequence:

Breaker, (CE 30): "Have completed a run from 4–8. I will see what numbers are available and what ones are requested before planning to keep the run or not."

Breaker, (CE 80): "I now have the sequence 45678 I hope that it does not interfere with any other's and have anxiety in thinking what I will do if someone needs one of mine."

Nonbreaker, (CE 100): "My part of the experiment is complete therefore I cannot trade."

Nonbreaker, (CE 100): "Finished my accomplishment. Satisfaction."

A similar contrast is provided by a selection of comments at Round 6, the first request round. None of these four *S*s had yielded to this initial request, but the comments and the CEs of those who were to break later indicate their awareness of the conflict.

Breaker, (CE 40): "Had number needed but giving up this would throw my sequence entirely—if, however, it is requested again I will prob. give it up."

Breaker, (CE 50): "7 has been requested don't know whether I should give it up and try for a new sequence."

Nonbreaker, (CE 100): "I couldn't help request section without ruining my sequence. Thus 100 and reason I passed."

Nonbreaker, (CE 100): "I hope the other boys are doing as well as I."

To return to the central problem of the study: what is the nature of the psychological process underlying yielding or not yielding to group pressure? The certainty estimate data lend support to the hypothesis that an important part of the behavioral variability in this situation can be attributed to *S*s' differential perceptual organization of the situation. The previous interpretation based on the response to competing demands must be modified to take account of *S*s' relative propensity to organize complex situations in such a way as to minimize the internal conflict.

SUMMARY

This study examined the relationship between *S*s' expressed level of certainty during an experiment and their observable behavior. Seventy-two undergraduates, randomly as-

signed to groups of six, were subjected to a quasi group-interaction situation.

Ss were led to believe that they were exchanging numbered counters with one another, through a messenger, to the end that all six members could achieve simultaneous solutions to a problem. Actually they were dealing only with *E* who throughout the task maintained a standard pattern of offers and requests for all. The crucial determination was whether or not *S* would break his own solution to yield to a simulated request from another group member for one of the numbers he was using.

At each transaction with the "messenger," *Ss* were required to estimate their level of certainty on a scale from 0 to 100. Certainty estimates at selected early junctures were then examined in the light of whether or not the *Ss* had ultimately yielded.

At the point when all *Ss* achieved their individual solutions and before any request had been made of them, some *Ss* expressed 100% certainty. These *Ss* later proved to be less likely to yield to the critical request than were those *Ss* who had expressed some uncertainty. This evidence is used to support an interpretation of yielding to group pressure which takes account of *Ss*' perceptual modes of organizing complex, ambiguous situations.

REFERENCES

ASCH, S. E. *Social psychology*. New York: Prentice-Hall, 1952.

BOOMER, D. S. Some determinants of the response to a conflict between personal and group goals: A methodological investigation of the quasi-group. Unpublished doctoral dissertation, Univer. of Calif., 1954.

CRUTCHFIELD, R. Assessment of persons through a quasi group-interaction technique. *J. abnorm soc. Psychol.*, 1951, **46**, 577–588.

Received January 8, 1958.

THE EFFECTS ON OPINIONS OF A CHANGE IN SCALE JUDGMENTS[1]

WALTER WEISS

Boston University

A PERSON'S position on a particular issue is manifested in part by the distribution of opinion items that he accepts. If the items are ordered on the issue dimension, his range of acceptance comprises an ordered portion of the total continuum. That is, the items with which one person agrees may range from extremely pro to mildly pro, while a second may accept items from extremely pro to neutral, and a third, items from mildly pro to mildly con, etc. Persons' ranges of acceptance may thus differ with respect not only to extent, but also to the location on the continuum of the pro and con limits beyond which items are not accepted. Obviously, many factors may influence the location and extent of a person's range of acceptance. As long as coercion is not involved, however, the selection of acceptable items should depend *in part* on the person's judgments of the relative locations of the items on the issue continuum and their relationship to his felt position on the issue. Thus, the scale of judgment that he focuses on the issue affects his inclination to accept or reject particular items. Consequently, an effective change in expressed opinions should result from judgmental changes that affect the perceived congruence of an item with his general point of view on the issue. If a person with a moderately pro position accepts items extending from neutral to strongly pro, for example, a change limited to his perception of the moderately con items, in which they come to appear extremely con, should not affect his expressed position. But a change that makes the moderately con items appear less extreme or even neutral in his perception may incline him to agree with them. The primary purpose of this study was a test of the effects of such a change in judgments on expressed opinions.

METHOD

An after-only arrangement involving three treatment groups was employed. The Wang-Thurstone scale of attitude toward the treatment of the criminal was modified to refer to delinquents, by substituting the word "delinquent" for "criminal" and making some minor changes in the wording of a few items.[2] One group, to be referred to as the "no standards" group, judged each of the 20 items of Form A of the modified scale on a 1–11 basis, in terms of its degree of approval or disapproval toward the use of punishment. A second group, to be referred to as the "end standards" group, made the same judgments but were provided with the following comparison standards for the two extreme categories (1 and 11): "Delinquents should only be treated with sympathetic understanding, guidance, and re-education; for the punishment of misguided youngsters is brutal and degrading"; and "All delinquents should be given long-term prison sentences and very severe physical punishment." Under this condition, it was expected that assimilation toward the end standards would occur for nearby items (Hovland, Harvey, & Sherif, 1957; Sherif, Taub, & Hovland, 1958) and perhaps contrast for the others, with the primary net effect being diminished discriminability among items in the middle region of the scale. The third group, to be referred to as the "middle standard" group, was provided with the following statement to be used as the standard for the midpoint of the scale: "Some delinquents should be punished; others should be given another chance." Although it was uncertain what effects an explicit middle anchor would have on judgments of opinion statements, the possibility of a resultant increase in central tendency led to the constitution of such a group. After making their judgments, all *S*s responded to the modified Form B in terms of their own opinions on the treatment of delinquents. The two categories of response were agree and disagree. A few general judgmental items completed the questionnaire.

The instructions stipulated that a rating of 1 was to be used for an item that seemed to express a very unfavorable opinion concerning the use of punishment, a rating of 11 for one indicating a very favorable opinion, and a 6 if an item seemed neither predominantly favorable nor unfavorable but appeared to signify neutrality or doubt concerning the use of punishment. The remaining numbers were to be used for statements that fell between either extreme and the neutral point of 6. The *S*s were informed that each of the 11 numbers need not be used, for they may feel that none of the 20 statements is a reasonable example of a particular position. They were also cautioned against permitting their own opinions on the topic to influence their judgments of the items. The explicit standards were presented only once, just before the first item in Form A.

[1] The research reported in this article was completed under Grant M-1351 from the National Institute of Mental Health.

[2] The need for equivalent scales and the knowledge that college students generally adopt a position opposing the use of punishment in the treatment of delinquents led to the selection of this topic.

TABLE 1
MEAN AFTER OPINIONS ON DELINQUENCY

	No Standards	End Standards	Middle Standard
Mean	4.29	4.68	5.22
SD	1.70	2.15	1.99
N	68	59	58

Booklets covering the three treatments were arranged in a random sequence, following a table of random numbers. They were handed out individually to the *Ss* who were seen during regular class periods at Boston University and American International College.[3] The purpose of the study was stated to be an examination of people's judgments of the opinions of others on a controversial topic.

RESULTS AND DISCUSSION

The opinion data were initially analyzed in terms of a 0, 1 system of scoring.[4] A score of 1 was assigned for a response which, in terms of the sense of the item, signified some favorableness toward the use of punishment, and a 0 to a response opposing punishment or favoring guidance and leniency in the treatment of delinquents. Neutral items could not be scored under this system. Accordingly, the opinion positions of the *Ss* are based on only 17 items of Form B and can range from 0 to 17. Nine of the items were phrased punitively and eight nonpunitively. The higher the score, the more punitive is the opinion position.

Effect of Standards on Acceptance of Items

The mean after opinion of each group is presented in Table 1. The F for groups, based on analysis of variance, has a p value of .05.[5] By t tests, only the middle standard group was found to be significantly more punitive than the no standards group ($p < .01$). In view of the method of scoring, this difference might be due to a decrease in the number of nonpunitive items accepted by the middle standard *Ss*, an increase in the number of punitive items accepted, or both. When the relevant data are examined, no significant differences are found among the groups in acceptance of the eight nonpunitively phrased items. But the mean number of punitively phrased items that the middle standard group accepted (2.36) is significantly higher than that accepted by the end standards group (1.88) or the no standards group (1.74). The p values of the between groups F and the individual t tests were all $<.05$.

TABLE 2
PERCENTAGE OF *Ss* AGREEING WITH EACH OF THE TWENTY ITEMS OF FORM B

Item	No Standards	End Standards	Middle Standard
6	2.9	6.8	0.0
12	14.7	16.9	20.7
17	73.5	72.9	70.7
2	92.5	96.6	94.7
8	95.6	91.5	86.2
20	82.4	84.7	86.2
5	75.0	69.5	74.1
11	95.6	84.2	87.9
16	94.1	91.5	98.3
14	76.5	74.6	79.3
1	22.1	30.5	36.2
19	92.6	93.2	94.8
4	80.9	76.3	87.9
10	35.3	37.3	44.8
15	10.3	11.9	22.4
7	7.4	6.8	15.5
13	16.2	16.9	27.6
18	4.4	10.2	10.4
3	0.0	1.7	8.6
9	0.0	1.7	6.9

These results are supported by an examination of the patterns of group agreement with each of the 20 items of Form B. Table 2 contains the percentage of *Ss* in each group agreeing with each item. The items have been arranged according to the original Wang-Thurstone scale values, with the first item the most strongly opposed to punishment and the last the most favorable. Although there are no significant differences among the groups on any item, a higher percentage of the middle standard group than of each of the others agrees with the last 12 items. Items 16, 14, and 1 fall in the neutral category, but the remaining nine are on the punitive side of the midpoint. Considering the percentages as repeated measurements and applying an arc sine transformation to them, an analysis of variance was performed involving groups, items, and groups by items interaction as the three sources of variation. The F for groups has a p value of .01. Individual t tests indicate that the middle

[3] The author wishes to thank Malcolm Klein of Boston University and Richard Sprinthall of American International College for distributing the booklets to their respective classes.

[4] The change in attitude object as well as the lapse of time between the development of the test and its current use precludes the use of the original scale values other than as an index of order.

[5] All p values are two-tailed.

standard group has a significantly higher average percentage of agreement in comparison with each of the others ($p < .05$). The first eight items of Table 2 were grouped as nonpunitive ones (i.e., their phrasing was such that acceptance indicated opposition to punishment or approval of guidance and leniency) and the last nine as punitive ones. Separate analyses of variance revealed a significant F for groups only for the punitive items. It is the middle standard group which is significantly different from each of the others; t test p values are $<.01$.

Thus, the analyses of the opinion data indicate that experience with the given middle standard led to the expression of a more punitive position. This effect was based essentially on differences in acceptance of the punitive items.

Effect of Standards on Variance and Extremity of Judgments

In terms of the theoretical suppositions underlying the research, group differences in expressed opinions should be associated with a particular pattern of judgmental differences. In general, the variance of scale judgments given by the middle standard group should be significantly less than that obtained under the other conditions. But more importantly, the middle standard group should judge the punitive items less extremely than do the other groups.

Based on the standard deviation of each person's scale judgments, the mean standard deviation was computed for each experimental group. Table 3 contains these means. By analysis of variance, the F for groups has a p value $<.05$. By t tests, the mean standard deviation of the end standards group is significantly greater than those of the other groups ($p = .05$ and $<.02$). Although the middle and no standards groups do not differ significantly from each other, the difference between them is in the required direction.

Examination of the distributions of median group judgments presented in Table 4 confirms this finding. As can be seen, the range of scale values is greater for the end standards group (1.9–11.0) than for the no standards group (2.7–9.3), which in turn has a slightly greater range than that for the middle standard group (3.1–9.1).[6] It is apparent from the data that besides differences in range, the three groups exhibit different distributions of median judgments. Most of the median judgments of the end standards group fall in the three extreme categories at each end of the scale; the midpoint category contains the majority of the median judgments of the middle standard group; and for the no standards group, the moderate categories are the locations of the majority of median scale values. These differences were found to be significant when analyses were made of the frequency data in Table 5. The p value of χ^2 for the entire table is $<.005$; each group is significantly different from each of the other two groups with p values ranging from $<.05$ to $<.005$.[7]

TABLE 3

MEAN STANDARD DEVIATIONS OF SCALE JUDGMENTS

	No Standards	End Standards	Middle Standard
Mean	3.24	3.54	3.13
SD	0.75	0.76	0.87
N	66	59	58

TABLE 4

MEDIAN SCALE JUDGMENTS OF EACH ITEM OF FORM A

No Standards		End Standards		Middle Standard	
Item	Median	Item	Median	Item	Median
17	2.7	8	1.9	8	3.1
8	2.9	12	2.0	5	3.1
2	3.6	2	2.2	2	4.0
20	4.2	17	2.4	11	4.6
12	4.8	6	2.5	12	6.0
5	5.3	5	3.1	16	6.0
16	5.3	11	4.3	17	6.2
6	5.5	20	4.9	6	6.4
11	5.5	16	6.2	20	6.4
1	6.4	1	6.4	1	6.4
14	6.4	14	6.4	14	6.4
19	6.9	19	7.1	3	6.7
3	6.9	15	8.8	19	6.8
13	7.9	4	9.3	18	6.8
18	8.5	10	9.5	9	6.8
15	8.7	7	9.6	15	6.9
10	8.9	13	9.9	10	7.1
7	8.9	18	10.1	7	7.1
9	9.2	3	10.5	13	8.2
4	9.3	9	11.0	4	9.1

[6] The Spearman rhos between the rank order based on the original Wang-Thurstone values for attitude toward the criminal and the order derived from the median judgments for each group are: end standards, .96; no standards, .84; and middle standard, .73.

[7] In terms of median scale judgments, none of the items whose original scale values would place them on

TABLE 5
FREQUENCY DISTRIBUTIONS OF MEDIAN JUDGMENTS OF FORM A ITEMS

Scale Categories	No Standards	End Standards	Middle Standard
1–3 or 9–11	5	13	3
4–5 or 7–8	11	4	5
6	4	3	12

TABLE 6
MEAN FREQUENCY OF USE OF SCALE VALUES IN JUDGING FORM A ITEMS

Scale Values	No Standards	End Standards	Middle Standard
1–3	6.38	6.95	6.22
4–5	1.55	1.14	1.24
6	4.83	4.24	5.95
7–8	2.15	1.88	2.28
9–11	5.08	5.78	4.31

When comparisons among the groups are made in terms of mean scale judgments of the nine punitively phrased and eight nonpunitively phrased items, significant F values for groups are obtained on each type of item (p values $<.05$). The means for the end standards groups were more extreme (i.e., smaller for the nonpunitive items and larger for the punitive ones) than were the comparable ones for the no standards group, whose means were more extreme than were those of the middle standard group.

Evidence for an Assimilation Effect

The data on judgments indicate that standards falling within or near the limits of a range of stimuli exert an assimilation effect. That is, judgments of items near to the standards shift in the direction of the standards. Although the actual scale values of the standards relative to those of the judged items were not obtained, over 80% of the groups provided with the standards agreed, in response to a multiple choice question, that they validly represented the designated scale categories.

To obtain some information on the empirical basis for the assimilation effect, the frequency with which each scale category was used was examined. For it may be supposed that assimilation, when it occurs, is due simply to an increase in the use of the particular category values given to the standards.

When analyses were made of the frequencies with which each scale value was used, only the neutral value of 6 yielded a significant F for groups ($p < .05$). The middle standard group used this category significantly more times than did each of the other two groups. Although differences in the use of the other categories were not reliable, there were certain obvious trends to the direction of the differences. To exhibit these, the data were organized in the form of Table 6. Besides the significant F for Category 6, the F for the combined scale values 9–11 had a p value $<.05$; by t tests, the middle standard group used these values significantly less frequently than did the end standards group. The end standards and no standards groups did not differ significantly from each other in the frequency with which they used any single category or combination of categories presented in Table 6. There is, however, an indication that the two sets of extreme categories (1–3 and 9–11) are employed more frequently by the end standards than the no standards groups. When the mean frequencies of use of these six categories are examined, the F for groups is significant ($p < .05$) and the mean of the end standards group is significantly greater than that of the no standards group ($p = .05$).[8]

Although numerous comparisons have been made, the data seem to reveal patterns of use of the scale values that are comparable to the distributions of median values for the 20 items of Form A (see Table 4 and 5). On the basis of these data, the assimilation effect does not seem to result *solely* from an increase in

one or another side of the midpoint was transposed by any of the groups to the other side of the midpoint. These results are consistent with those obtained by Fehrer (1952). As she noted, when judgments are made of opinion items along a commonly used continuum, gross directional cues can be derived from the customary meanings and usages of language. For example, the directional cue afforded by the phrase "must be punished" in the statement "delinquents must be punished" is so unambiguous that transposition to the lenient side of the midpoint is not likely to occur.

[8] The mean number of categories used by each group in judging all 20 items was: end standards, 6.25; no standards, 6.48; and middle standard, 5.64. By analysis of variance, the F for groups was not significant. Thus, the experimental variations affected the distributions of judgments and not the number of categories used in making the judgments.

the frequency with which the scale values of the designated standards are used. Such an increase held under the middle standard condition. But under the end standards condition, although some increase occurred in the use of scale values 1 and 11, it was only when the six end categories were combined that a significant effect was exhibited.

The Relationship Between Judgments and Opinions

In general, the data on group differences in expressed positions and in patterns of scale judgments provide empirical support for the theoretical assumptions underlying the research. A marked reduction in discrimination among opinion items around the more punitive limit of the range of acceptance apparently led to an extension of the range in that direction. Since an overwhelming majority of the *S*s in all groups agreed with at least six of the eight nonpunitive items, judgmental effects on these would not be expected to influence agreement with them. The assertion of support for the research hypothesis is based also on the assumption that the effects of the standards perseverated and influenced in a comparable manner judgments of the Form B statements. Considering the relative equivalency of the forms and the very brief delay between responses to them, this assumption seems tenable.

Before data are presented concerning the within-groups relationship between judgments and opinions, some assumptions must be made explicit. It is assumed that judging the 20 items of Form A in the absence of explicit standards had little, if any, effect on opinion responses. It is likely that the *S*s' opinions were already crystallized as a result of prior exposure to information and points of view expressed in the mass media, in relevant college courses, and by other people. The topic of juvenile delinquency has been a relatively "live" one over the past few years. It is also assumed that the opinions expressed by the end standards group were not affected by the preceding judgment experience; the difference in mean opinion position between this group and the no standards one did not approach significance. Since the judgmental experience with the middle standard did affect expressed opinions, the relationship between judgment and opinion for this group is not readily interpretable and, therefore, is not included in this analysis.

TABLE 7

RELATIONSHIP BETWEEN *S*s' OWN OPINIONS AND DISTRIBUTIONS OF MEDIAN JUDGMENTS OF FORM A ITEMS

Scale Categories	No Standards		End Standards	
	≥5[a]	<5	≥5	<5
1–3 or 9–11	3	12	6	14
4–5 or 7–8	11	5	8	3
6	6	3	6	3
p_z[b]	<.02		<.05	

[a] Opinion scores based on responses to Form B.

On the basis of the 0, 1 system of scoring the opinion responses, *S*s were divided into two subgroups: those with an opinion score of 5 or greater (moderate opposition to punishment) and those with a score less than five (strong opposition to punishment). (This cutoff point is slightly above the over-all median of 4.72.) The judgmental patterns of these subgroups were then determined. The data in Table 7 are based on the median scale judgments for the 20 items of Form A. As can be seen, the strongly opposed subgroups place significantly more items in the end categories than do the moderately opposed subgroups. (Approximately equal numbers of items were placed by the subgroups in the 1–3 as in the 9–11 categories.) Rank order correlations were computed between the rankings assigned the items of Form A by the opinion subgroups of the same experimental group. For the end standards subgroups, the correlation is .87 and for the no standards subgroups, .53. Thus, the ordinal arrangement of the items appears to have been affected more by the *S*s' opinions when no standard was present than when the end standards were given.

If a person's opinion position on an issue is crystallized and salient, it may function to some extent as an implicit standard and thereby affect his judgments of issue-related stimuli. (In view of the publicity given to the general problem of delinquency in recent years, it is likely that the two required conditions of crystallization and saliency obtained for most *S*s.) Since the opinion position of one subgroup is close to the lenient end of the scale and that of the other is nearer to the midpoint, the

distributions of judgments for these subgroups should differ primarily in the number of items located around the middle region of the scale. That is, the distributions should approximate what might be expected if explicit anchors were located on the scale at the mean positions of the subgroups. Furthermore, if opinion position does function as an implicit standard, its effects on judgments should be less when explicit standards are given than when they are absent. This expectation accords with the finding of a lower rank order correlation between the opinion subgroups under the no standards condition than under the end standards one. Also, if it is assumed that the "impressiveness" and saliency of an implicit standard is greater for an extreme opinion position than for a moderate one, then the following prediction can be made: the effectiveness of explicit standards in producing judgmental differences will be less when the comparison groups hold to the same extreme position than when they hold to the same moderate position. The current data provide some support for this inference. The rank order correlation between the median judgments of the two strongly opposed subgroups is .95, but only .68 when the moderately opposed subgroups are compared. In general, the data indicating a within-group relationship between opinions and judgments are consistent with the results of some previous research (Cartwright, 1941; Hovland & Sherif, 1952; Sherif & Hovland, 1953).

SUMMARY

The primary purpose of the research was to examine the effects of changes in the judgments of the scale positions of opinion items on the judges' own opinions on the issue. Three groups were formed by random assignment of subjects. One (the "no standards" group) judged on a 1–11 basis the scale positions of a modification of the Form A items of the Wang-Thurstone scale of attitude toward the treatment of the criminal. The modification involved the substitution of the word "delinquent" for "criminal." The *S*s then responded to the similarly modified Form B in terms of their own opinions. Another group (the "end standards" group) responded similarly but were provided with explicit statements to be used as reference standards for the 1 and 11 categories when judging the Form A items. The third group (the "middle standard" group) differed from the others only in that they were given an explicit reference statement for the midpoint category of 6. Some additional judgmental questions completed the questionnaire.

The results indicate that the explicit middle standard enhanced central tendency and the end standards counteracted it. The mean opinion position of the middle standard group was significantly different from that of the no standards group; but the position of the latter group was not different from that of the end standards group. Thus, a certain kind of change in judgments was found to be associated with a change in opinion position. The differences among the opinion positions of the groups were shown to be due primarily to an increase in acceptance of punitive items by the middle standard group. Under the end standards and no standards conditions, opinions on the issue were found to be related to judgments of the scale positions of the Form A items.

REFERENCES

CARTWRIGHT, D. C. Relation of decision-time to the categories of response. *Amer. J. Psychol.*, 1941, **54**, 174–196.

FEHRER, ELIZABETH. Shifts in scale values of attitude statements as a function of the composition of the scale. *J. exp. Psychol.*, 1952, **44**, 179–188.

HOVLAND, C. I., HARVEY, O. F., & SHERIF, M. Assimilation and contrast effects in reactions to communication and attitude change. *J. abnorm. soc. Psychol.*, 1957, **55**, 244–252.

HOVLAND, C. I., & SHERIF, M. Judgmental phenomena and scales of attitude measurement: Item displacement in Thurstone scales. *J. abnorm. soc. Psychol.*, 1952, **47**, 822–832.

SHERIF, M., & HOVLAND, C. I. Judgmental phenomena and scales of attitude measurement: Placement of items with individual choice of number of categories. *J. abnorm. soc. Psychol.*, 1953, **48**, 135–141.

SHERIF, M., TAUB, D., & HOVLAND, C. I. Assimilation and contrast effects of anchoring stimuli on judgments. *J. exp. Psychol.*, 1958, **55**, 150–155.

Received January 16, 1958.

STRUCTURAL ANALYSIS OF DREAMS:

CLUES TO PERCEPTUAL STYLE[1]

BERNICE T. EIDUSON

Hacker Foundation for Psychiatric Research and Education

RECENT experimental work (Klein, 1951) has supported the hypothesis that general cognitive styles or organizing principles of perceiving characterize each individual. These cognitive system-principles seem to operate as personality constants and, in the ways they select, organize, and distort perception, reveal character trends that have emerged in the person's development. Many experiments in this tradition (cf. Adams, 1957; Miller, 1951) have studied perception under subliminal conditions, or in situations where the drive-state has been very intensified, or where fatigue or drugs have weakened integrated functioning. It has been assumed in these experiments that perception occurs simultaneously on many levels of awareness. The question arises whether the same cognitive styles that have been described mainly under normal waking cognition are found on different levels of consciousness. It is to this question that the present investigation was directed.

This study proposed to test whether persons who show differences in the perceptual attitude of flexibility–rigidity as estimated from the Rorschach test also differ predictably along this dimension in the perceptual response pattern of their dreams. The dimension of flexibility refers to the ability to vary, shift, and alter the response pattern. It implies an openness to change, a tolerance for free fantasy, and for lability in emotion, a tendency toward diversity in ways of expression. Its polar opposite is rigidity, which reflects a tight defense against variation or change, a reluctance in experimenting with the new, and a complacency with the original perceptual response. Empirical studies by Klein (1951) and others (Klein & Schlesinger, 1951) have yielded increasingly impressive indications that this dimension comprises one of the regulatory principles in cognition. Our own preliminary work has suggested that the flexibility–rigidity dimension was appropriate to performances both on the Rorschach test and in dreams.

The Rorschach and dream situations were selected for perceptual study for two reasons: first, the states of psychological awareness that they represent are different—the dream reflects functioning when defensive and integrating mechanisms are relaxed, with the usual spatiotemporal, reality-oriented cues abandoned, except as they are instituted in reporting the dreams; the Rorschach occurs in the waking state and expresses the highly individualistic primary modes of perception only as they are subordinated to the logical, rational framework that the test situation demands. Secondly, both are perceptual situations that occur naturally in the clinical therapeutic experience. We were interested in seeing whether clinical material could be felicitously used in perceptual studies and perhaps provide a source of data not normally available in the laboratory situation.

METHOD

Rorschach Criteria

Rorschach records of 18 outpatients who were being seen in intensive treatment at a psychoanalytically oriented clinic were reviewed by three clinical psychologists working independently. The Rorschachs had been administered as part of the diagnostic study prior to treatment and independent of this study. The psychologists were instructed to consider the test material globally and to rank each subject in terms of the hypothesized perceptual variable, using the following a priori criteria: number of major determinants used simultaneously, amount of spread in use of possible determinants for each response, variety of major determinants used in the first response, variability in content categories, variability in sequence, degree of elaboration of responses, degree of constriction of record, number of responses to each card. Judges were free to decide on the relative importance of these criteria in each case. The coefficient of concordance, corrected for continuity, was applied to provide the estimate of agreement among judges. The F ratio obtained was 9.5; the W' was .76, which indicated that the three sets of ranks showed agreement beyond what might be produced by sampling variance. A composite ranking for the group was made based on the sum of ranks. From the ranking of the total group, the six sub-

[1] Presented in part at the XVth International Congress of Psychology, Brussels, 1957.

jects at each end of the ranking were selected to form the population for this investigation.

Subjects

The total group was composed of four men and eight women ranging in age from 23 to 37. Two men and four women were rated as "flexibles," and their range in age and educational background was roughly similar to the six "rigids." All had been diagnosed as presenting neurotic or character disorders. All of the *S*s had been in treatment from 6 to 18 months prior to the start of this investigation. They were known to be persons who were so intensively involved in treatment that they were producing dreams regularly during their therapeutic hours. The patients were unaware that they were *S*s, and there has been no indication that the process of their psychotherapy was affected in any untoward way by their being used as research *S*s without their knowledge.

Dream Protocols

Psychotherapists were instructed to record verbatim all dreams produced during course of treatment in the five-week period of this study. Since the psychotherapists had been accustomed to collecting dreams as part of their regular therapeutic procedure, little variation in technique was required. Some therapists recorded in their notebooks as customarily done; others used forms provided by the investigator. All dream material occurring during one night was considered as a single dream unless the *S* spontaneously mentioned awakening between dreams. A mean of 13 dreams per *S* (145 dreams in all) were collected.

New techniques of analysis had to be adopted for treatment of the dream material for the present purpose. Classically dream analysis (Freud, 1950; Hall, 1951; Leavitt, 1957) has dealt only with the content of imagery and its expression through the mechanisms of displacement and condensation. The concrete image was taken as the scorable response in order to avoid any interpretation of content. Previous study of sensory modalities employed in dreams (Freud, 1950; Knapp, 1956) had pointed preponderantly to their visual character. In our work, however, we were interested in the readiness for fluctuation, shift, and variability as it was revealed in the sense imagery.

The concrete dream imagery was rated on the following categories that were presented on check list forms to psychotherapists:

a. Sense modalities employed—(visual, auditory, tactual, kinesthetic, olfactory)

b. Degree of intensity of imagery—(rated on a five-point scale from "very intense or vivid" to "faint")

c. Amount of affect accompanying imagery—(three-point scale: none or minimal; some or moderate; great or strong)

d. Kind of affect accompanying imagery—(five-point scale varying from "very pleasant" to "very unpleasant")

Psychotherapists were given criteria for rating of dream imagery and were given pretraining. At no time, however, were the psychotherapists informed about the perceptual dimension under investigation. The verbatim recording of dreams permitted independent rating of raw data by the investigator, and all discrepancies of more than one scale point were discussed and re-evaluated with the therapist. On a group of 20 dreams, the interrater agreement (psychotherapist and investigator) was .86.

Each dream was rated according to the categories listed above, and then the total set of dreams for each *S* tabulated. Totals were converted to percentages to allow for difference in the number of dreams reported by *S*s during the period of investigation. Then, to provide an estimate of flexibility or rigidity in dream imagery, the percentages were analyzed in terms of the following a priori criteria:

A. Total number of sense modes used simultaneously. (Percentage of times 1, 2, 3 modes used together in the *S*'s series of dreams)

B. Incidence of variance in primary mode in which imagery was expressed. (Percentage of times primary mode was visual; percentage of times primary mode, auditory; etc.)

C. Spread of modes when visual mode was primary one used. (Percentage of times visual was accompanied by auditory; by tactual; by auditory and tactual; etc.)

D. Variation in intensity with which images were expressed. (Percentage of times there was a change in degree of intensity from that which *S* used most frequently, the modal response)

E. Variation in amount of affect attached to imagery. (Change from average [mode])

F. Variation in kind of affect attached to imagery. (Change from average [mode])

RESULTS

The experimental hypothesis was tested by correlating each *S*'s rank on the flexibility–rigidity dimension in the dream imagery with his rank on the dimension as evaluated from his Rorschach test performance. In order to obtain the *S*s' ranks on dream imagery, first the nonparametric sum of ranks was applied to each of the above criteria to test the significance of the difference between means of *S*s who had been grouped as "rigid" or "flexible" on the basis of their Rorschachs. The results showed all but one of the variables to be significant (Table 1). Therefore, *S*s who had been rated as "flexible" on their Rorschachs were generally rated as "flexible" when the categories for flexibility in dream imagery were considered individually. Then with the median scores of the total group of 12 *S*s as the cut-off points in each of the categories, each *S*'s performance was ranked in the five categories which had shown significance. The final rank ordering of *S*s on the flexibility–rigidity dimen-

sion in dream imagery was obtained by taking the average (mean) of these ranks.

A rank correlation coefficient of .74, significant at the .01 level, was obtained when the ratings of flexibility–rigidity in Rorschach test performances were compared with those found in dream imagery. Persons who are flexible in their mode of organizing Rorchach responses thus tend to be similarly flexible in their use of imagery in dreams.

Behavioral Correlates of Flexibility–Rigidity in Dreams. Confirmatory validity for the rating on the flexibility–rigidity dimension in dreams was sought by evaluating whether the "flexible" *S*s actually functioned in a flexible way in various aspects of their behavior and whether the "rigid" *S*s displayed rigidity in their functioning. Psychotherapists who knew their behavior patterns intimately were asked to rate them in these areas: work, interpersonal relationships, attitudes toward self, imaginative and ideational resourcefulness, and emotional behavior. A four-point scale in which the degrees of flexibility were distinguished was used. Psychotherapists were instructed that flexibility referred to the extent to which the *S* showed variation, instabilities, shifts, or unsteadinesses in his behavior pattern. They were also told that we were interested primarily in the amount of shift or change in behavior rather than in what this variability meant in the patient's personality picture and therefore that a high degree of flexibility could for some patients signify the potential for appropriate and varied responses to different situations, while for others it might mean the lack of directiveness or the unpredictability in behavioral response. This part of the study was done four weeks after the dream collection had been discontinued and was the first time psychotherapists were informed about the perceptual dimension under study.

To obtain a composite rating of "flexible" or "rigid" in behavior for each *S*, the clinical ratings made in each of the five specific areas were averaged, each having equal weight (Table 2). The mean of the four scalar points (2.5) served as the cutting-off point for the general rating as "flexible" or "rigid" (cf. Table 2.). Using Fisher's exact test on 2 × 2 contingency tables, the independence of ratings of actual function and of dream imagery was tested (Finney, 1948).

TABLE 1
RANK ORDER RATINGS OF "FLEXIBLES" AND "RIGIDS" ON DREAM IMAGERY VARIABLES

Subject Number	Variables[a]					
	A*	B**	C**	D	E*	F*
"Rigid" subjects						
2	6	10	9	7	7.5	2
3	12	8	9	2	12	8
6	9	12	12	12	7.5	11.5
7	6	7	11	11	9	11.5
8	10	11	9	5	5	10
11	11	9	7	9	11	9
"Flexible" subjects						
1	6	4.5	5	1	4	4
4	1	1	1	3.5	3	1
5	2	6	6	10	10	3
9	8	4.5	3	8	2	5.5
10	4	2	4	3.5	1	7
12	3	3	2	6	6	5.5

[a] Variable A—Total number of sense modes used simultaneously; B—Incidence of variance in primary mode in which imagery was expressed; C—Spread of modes when visual mode was primary one used; D—Variation in intensity with which images were expressed; E—Variation in amount of affect attached to imagery; F—Variation in kind of affect attached to imagery.

*.05 critical point for smaller rank totals for $n_1 = 6$, $n_2 = 6 = 26$ (White, 1952; Wilcoxon, 1945), computed on only one side.

** .01 critical point for smaller rank totals for $n_1 = 6$, $n_2 = 6 = 23$ (White, 1952; Wilcoxon, 1945), computed on only one side.

The correlation between dream imagery and overt behavioral functioning was significant at the .05 level. This would confirm the hypothesis that flexibility operates with a high degree of consistency in the *S*'s style of perceiving and in his everyday functioning. It suggests that such organizing modes may act as "sets" for behavior. This relationship between dreams and behavioral functioning also assures us that an appropriate measure was developed for reflecting the flexibility–rigidity dimension in dreams.

With respect to psychotherapists' ratings for specific areas, significant lack of independence was found between ratings of flexibility in dream imagery and in attitude toward self. This was the only area that reached the .05 level of significance. However, ratings in areas of interpersonal relationships, emotional behavior, and ideational resourcefulness were significant at the .10 level (when chi square

TABLE 2
RATINGS OF FLEXIBILITY–RIGIDITY IN BEHAVIOR

Subjects	Variables					
	Work	Interpersonal Relationships	Attitudes Toward Self	Ideational Resourcefulness	Emotional Behavior	General Rating on Flexibility Dimension[b]
1	4	3	4	3	4	*f*
2	3	1	3	1	1	*r*
3	1	1	1	1	1	*r*
4	—[a]	3	4	2	3	*f*
5	3	3	3	3	4	*f*
6	2	2	3	3	3	*f*
7	2	4	2	4	4	*f*
8	4	2	2	2	2	*r*
9	4	4	4	3	4	*f*
10	2	4	3	3	3	*f*
11	3	2	2	2	2	*r*
12	1	3	4	3	2	*f*
	$\chi^2y = .12$ $p > .50$	$\chi^2y = 2.83$ $p < .10$	$\chi^2y = 5.18$ $p < .05$	$\chi^2y = 2.83$ $p < .10$	$\chi^2y = 2.83$ $p < .10$	$\chi^2y = 5.18$ $p < .05$

[a] No work history.
[b] Obtained by getting mean of ratings listed and using mean of scalar points (2.5) as cut-off point.

with a Yates correction was applied), suggesting a tendency toward consistency in these specific areas of behavior and imagery (cf. Table 2).

DISCUSSION

In this study the same cognitive style principle has been shown in behavior on different levels of psychological awareness. In the dream where ego controls are relaxed and the primary process takes dominance, the flexibility–rigidity style principle was found to operate as a mode of organizing perception, as it does in the cognitive test situation administered under normal waking conditions. In the study of overt behavioral functioning the same cognitive style principle was also observed. Because of the limitations presented by this circumscribed procedure, this can at best be a tentative conclusion demanding more refinement and validation. However, the results point to the need for exploring perception under a variety of differing conditions of awareness.

Speculation as to the significance of these findings directs attention to the possibility that precursors of cognitive style patterns that we see in rational logical secondary processes may exist in primary process thinking. Classically, primary thought processes have been described as being mainly unorganized, but the question is whether or not the precipitates of the organization which ultimately the ego displays are contained therein. The early psychoanalytic position stated that the id, out of which the ego was derived, was completely chaotic and without order; however, psychoanalysis today holds that from the very first the original endowment shows primitive ego regulatory mechanisms, that there is no period in the individual's life when some suggestion of these mechanisms is not found, and that even in reflex activities these are already operative. This study suggests that the primitive ego formation is likely to have structural dimensions, and that in early development these may set the stage for the perceptual response pattern which we see in adult purposeful and directed function.

One further bit of clinical data is provocative. Among the material available on each *S* was the predominating symptom picture. Inspection of this showed that the symptoms of the *S*s in the flexible group were alike in that all were primarily directed out onto the environment: four showed impulsive behavior with sexual acting out or difficulties with authority, and two were paranoid character disorders.

The patients in the rigid group by contrast tended consistently to show symptoms which were directed inwardly, or toward themselves: two, psychosomatic or conversion disorders; three, depressions; one, phobias. The implica-

tions of this alignment are not clear but are suggestive of differences similar to those reported between direction of aggression and certain biochemical correlates—for example, in the work of Funkenstein, King, and Drolette (1954) and others (Eiduson, Crumpton, & Brill, 1957) who have related direction of aggression to the epinephrine-norepinephrine ratio. This relationship points to a fruitful avenue for further research, although the mere generalization of the correlates of hypothesized attitudes does not explain the processes involved.

SUMMARY

This experiment was directed to the question of whether the same perceptual style principles that have been described under normal waking conditions are found on different levels of consciousness. The experimental hypothesis was that persons who show differences in the perceptual attitude of flexibility–rigidity as estimated from the Rorschach test also differ predictably along this dimension in the perceptual response pattern of their dreams. Twelve patients in intensive treatment at a psychoanalytically oriented clinic were administered the Rorschach test, which was rated for degree of flexibility on a number of a priori criteria. From each of the subjects, who were producing dreams regularly during their therapeutic hours, a mean of 13 dreams (145 in all) was collected. Concrete dream imagery was analyzed for flexibility or rigidity according to a number of a priori criteria which were considered appropriate to the material.

Results confirmed the experimental hypothesis. Ratings as to the flexibility shown by subjects in five areas of overt behavior and in their functioning in general also bore out the rating of flexibility or rigidity displayed in dream imagery. The findings indicate that in psychological behavior at varying levels of awareness there is consistency in the operation of the cognitive style principle of flexibility–riddidity, and suggest the importance of further exploration of perception under different conditions of consciousness.

REFERENCES

ADAMS, J. K. Laboratory studies of behavior without awareness. *Psychol. Bull.* 1957, **54,** 383–405.

EIDUSON, S., CRUMPTON, E., & BRILL, N. Q. Plasma cathecol amines as related to direction of aggression in human beings. *Proc. 15th Int. Cong. Psychol., Brussels,* 1957.

FINNEY, D. J. The Fisher-Yates test of significance in 2 x 2 contingency tables. *Biometrika,* 1948, **35,** 145.

FREUD, S. *The interpretation of dreams.* A. A. Brill (Trans.) New York: Random House, 1950.

FUNKENSTEIN, D. H., KING, S. H., & DROLETTE, M. The direction of anger during a laboratory stress-inducing situation. *Psychosom. Med.,* 1954, **16,** 404.

HALL, C. S. What people dream about. *Sci. Amer.* 1951, **184,** 60.

KLEIN, G. S. The personal world through perception. In R. R. Blake & G. V. Ramsey (Eds.), *Perception: An approach to personality.* New York: Ronald, 1951.

KLEIN, G. S., & SCHLESINGER, H. J. Perceptual attitudes of "form-boundedness" and "form-lability" in the Rorschach. *J. Pers.* 1951, **19,** 289.

KNAPP, P. H. Sensory impressions in dreams. *Psychoanal. Quart.* 1956, **25,** 325.

LEAVITT, H. C. Teleological contributions of dreams to the waking ego. *Psychoanal. Rev.,* 1957, **44,** 212.

MILLER, J. G. Unconscious processes and perception. In R. R. Blake & G. V. Ramsey (Eds.), *Perception: An approach to personality.* New York: Ronald, 1951.

WHITE, C. The use of ranks in a test of significance for comparing two treatments. *Biometrics,* 1952, **8,** 33.

WILCOXON, S. Individual comparisons by ranking methods. *Biometrics Bull.* 1945, **1,** 80.

Received January 27, 1958.

AUTONOMIC RESPONSES AND THEIR RELATIONSHIP TO RACE ATTITUDES[1]

FRANK R. WESTIE AND MELVIN L. DE FLEUR

Indiana University

This is an experimental study of autonomic responses of prejudiced and unprejudiced persons viewing interracial situations. Changes in galvanic skin resistance, finger pulse, duration and amplitude of heart beat, and duration of heart cycle were recorded. The relationship of these measures to race attitudes as indicated both by scales and by open-ended interviews suggests that attitude-response syndromes include emotional tensions and affective feeling tones not under the individual's direct voluntary control or subject to his conscious examination. Experimental exploration of this more subtle dimension of attitude behavior may lead to a clearer understanding of the concept "attitude."

Most direct approaches to the assessment of attitudes rest upon at least two tacit assumptions: (*a*) that people can introspectively assess what their attitudes are (or "would be" under some hypothetical circumstance) toward the attitude object in question, and (*b*) assuming that people *do* know what their attitudes are, that they are both willing and able to communicate these faithfully to the investigator. This foundation of assumptions, upon which modern structured attitude scaling rests, stands in need of critical examination. We question whether all attitudinal phenomena exist for subjects on such a verbalized level of awareness as to permit them to look into themselves, and then say, in effect, "this is how unfavorable I am." If this aspect of a person's image of himself is unclear, there is scant possibility that it can be effectively communicated to the investigator, regardless of his ingenuity in constructing orderly scales. Furthermore, if attitudes include deep-seated emotional orientations, respondents may be incapable altogether of either perceiving or communicating such information. Measurement of such emotional orientations is not possible with traditional paper-and-pencil techniques.

This is not to say that structured scales and paper-and-pencil techniques are useless. On the contrary, we have used them extensively in the present research. The point is rather that they do not yield the whole picture of attitudinal behavior, but provide an image of only one dimension, the verbal.

We find it useful to conceive of attitudinal phenomena as a multidimensional complex, without any assumption of a one-to-one correspondence between the dimensions. One dimension is the emotional orientation we seek to examine; another is the verbalized behavior; still a third dimension includes overt approach-avoidance acts which to many seem to be the "real" behavior that attitude scales should predict. All of these are actual behaviors and constitute important aspects of the attitude complex. Such phenomena exist as recurrent behaviors, and the task of attitude research, as conceived here, is a search for consistency of response within these dimensions.[2]

Although existence of such dimly perceived feeling states is not difficult to demonstrate, their measurement proves to be a somewhat more exacting task. For example, many persons who have had experience in interviewing white people about their attitudes toward Negroes have felt a shortcoming in standard paper-and-pencil techniques. Individuals are

[1] We wish to acknowledge our indebtedness to the Social Science Research Council for financial assistance and to Rolland C. Davis of the Indiana University Psychology Department who gave generously of both his time and equipment in connection with the physiological measurements. The recording of physiological responses was partially supported by funds under contract No nr 908 (03) between the Office of Naval Research and the Indiana University Foundation.

[2] The reviews of the literature of the past thirty years by Allport (1935), Campbell (1950), and Green (1954) all stress that one characteristic is almost uniformly mentioned in attitude definitions: consistency of behavior, either in action toward the attitude object or in verbalizing about it. This usage, in effect, equates the term attitude to a probability (of recurrent behavior). Campbell's definition, as paraphrased by Green, expresses this focus: "An individual's social attitude is an enduring syndrome of response consistency with regard to a set of social objects. Many theorists more specifically include in their definitions some statement concerning "emotional overtones," "affective feeling states," or some similar phrase.

frequently encountered who are determined to be "big" about the "Negro question," not only to the interviewer but to themselves. The strictly verbal picture that emerges from such interviews is that of a person who is generally favorable in his attitudes toward Negroes. But more subtle responses speak eloquently of another picture—a certain uneasiness, the clenched fist, the restless stirring, the new atmosphere of forced friendliness—all of this following closely upon a pleasant, relaxed easygoing atmosphere built around a discussion of an affectively neutral topic. We have observed that this change occurs when the first question about Negroes is asked. Although the interviewer may be quite aware of these subtle cues, he ordinarily lacks techniques with which to control, quantify, or record such responses.

Considerations of this sort led the writers to embark upon a study of emotional tensions in situations of race relations, as indicated by autonomic responses. The relationship of such responses to measures of race attitudes obtained by more structured methods may allow meaningful exploration of this neglected dimension of attitudes.

The problem may be stated as follows:

1. Are there consistencies in autonomic response associated with exposure to objects toward which persons have (scale-measured) attitudes?

2. If there are measurable associations between autonomic responses and attitudinal involvement, what is the direction of the (scale-measured) attitudinal involvement? Do those persons most unfavorable in their verbal response to the attitude-object manifest more pronounced autonomic responses than those more favorable in their verbal responses? Or is the reverse true?

METHOD

The investigating involved three phases: the attitude testing program, the measurement of autonomic responses to attitude stimuli, and the postexperimental interview.

The Attitude-Testing Program

Subjects (*Ss*) were selected on the basis of scores on the Summated Differences scales.[3] These scales require the *S* to indicate the degree to which he is willing to associate (in a variety of social settings) with Negroes of a variety of socioeconomic levels as compared to whites of comparable socioeconomic levels. The total score, based on the summated differences between responses to whites and Negroes of the same socioeconomic level, provides an index of positive or negative verbal attitudes toward Negroes. This battery of scales, made up of some 500 items, was administered to 250 freshmen students at a large university. From this population, a control group of 23 *Ss* was matched in over-all frequency distribution with an experimental group of the same size in terms of age, sex, social class, residential history, and previous contact with Negroes. The experimental ("prejudiced") group was selected from students who scored in the upper 25% of the Summated Differences scale, while the control ("unprejudiced") group was drawn from those scoring in the lower 25% of the distribution. Half of each group were males, the other half females.

The Laboratory Study

In the laboratory study, the *Ss* of each group were exposed individually to a series of colored photographic slides depicting whites and Negroes in various situations. During these presentations the *S* was connected to apparatus designed to measure GSR (galvanic skin response), FP or finger pulse volume (measured by plethysmograph), amplitude and duration of heartbeat (measured by electrocardiograph), duration of heart cycle (time required for the heart to pass from one state to the same state again, as measured by cardiotachometer). The physiological measures required about an hour per *S*. The data were gathered during a 12-week period.

The colored slides were posed and photographed specifically for this study. They were projected through a port into a small air-conditioned room where the *S* sat relaxed in a chair. Each slide portrayed two "good-looking" young adults sitting side-by-side looking at each other with a pleasant, friendly expression. One slide was included for each possible combination of males and females, whites and Negroes. In addition to these pair slides, a set of four single-person slides was included, portraying a single member of each race and sex. The age of the person photographed approximated that of the *Ss*. The models were dressed in dignified apparel; the background behind the seated persons included a small table with a lamp and a partial view of a window with a drapery. The white and Negro models appearing in the slides were not known to the *Ss*. The slides were thus designed to avoid the arousal of strong autonomic reactions for reasons extraneous to the problem at hand.

A suitable rest and standardization period was allowed for each *S* after being attached to the apparatus. Each *S* was then shown the 12 stimulus slides. The four single-person slides were shown as a first set, followed by eight pair slides. For both the first set and the second set, an independently drawn random order of presentation was used for each *S* so as to control for serial effects. Each slide was shown for 30 seconds with a 30-second rest between each presentation.

Although five separate autonomic physiological

[3] For a detailed description of the manner in which these scales were constructed, their reliability coefficients, etc., see Westie (1953).

responses to these slides were recorded, the present paper considers only two of these: the GSR, recorded through electrodes placed on the palm, and finger pulse volume (FP), registered by an electrode at the end of the index finger. An index of change in GSR patterns was devised which states changes in skin resistance as a percentage of the resistance at the beginning of each stimulus presentation. Since the GSR activity was permanently recorded by a stylus on a recording drum, it was possible to synchronize such changes with the exact time of stimulus presentation. A similar system was used for recording FP changes. Amplified FP activity was permanently recorded on a separate drum paper, and synchronization of this record with the presentation of the stimuli was also possible.[4]

The Post-experimental Interview

Each *S* was interviewed immediately following the measurement of his physiological responses. During this hour-long session, which took place in an office far removed from the laboratory, the *S* responded to a variety of structured and open-ended questions intended to help in giving meaning to the physiological responses. Projective interpretations were elicited to slides comparable to those he viewed in the laboratory, and topics such as the *S*'s previous experiences with Negroes, his personal and social background, his reference groups, and his degree of willingness to be photographed with Negroes were explored.

RESULTS

Quantified variations in these physiological responses were arranged in a factorial analysis of variance design. The statistical analysis was complicated by the fact that each set of slides (single person slides vs. paired slides), and each physiological measure yielded separate data.

In addition, for the GSR, the continuous kymograph records provided several ways of obtaining measures of the results. Thus, GSR responses to the single-person slides were analyzed on the basis of measures made during the *first third* of the 30-second period during which each *S* viewed each slide. Similarly, the *last two thirds* of each stimulus presentation provided the basis of another complete analysis. Still another analysis was made of changes in autonomic activity during each 30-second stimulus period taken as a unit. From these various combinations, six analyses were completed.

In the present paper only the analysis of the last two thirds of the stimulus period is presented. Although the results of the different analyses are quite similar, our data seem to indicate that fullest response occurred during this last 20-second period of the stimulus presentation.[5]

Table 1 shows the general statistical design and gives the mean level of finger pulse (FP) response to the single-person slides by various categories of *S*s to the different types of slides. In addition, the usual summary table for the analysis of variance, based on the FP responses to the single-person slides, appears in Table 2. A verbal summary of the significant sources of variance in all unreported analyses is given in Table 5.

Examination of Table 1 indicates that FP activity was generally less for females than for males, but Table 2 reveals that the difference is not significant. Similarly, the data in Table 1 show that prejudiced *S*s had generally lower levels of FP response to the Negro slides than the unprejudiced *S*s, but, again, the difference is not significant. The reaction patterns of these two *S* groups to the white slides were very similar. However, statistically significant results are evident when both *S* sex and *S* attitude are taken into account jointly. This finding can be more clearly understood by com-

[4] Space does not permit discussion of details concerning the apparatus, the recording devices, the nature of the laboratory, methods of translating response records into quantitative indices, etc. The apparatus is described in detail in Davis, Siddons, and Stout (1954). The procedure employed in the quantification of the kymograph records was relatively simple. For the finger pulse records, the greatest amplitude of finger pulse volume for a 5-sec. period (2.5 to 7.5 sec. after onset of stimulus) was divided by the greatest amplitude of FP volume for a 5-sec. interval preceding the onset of stimulus.

The GSR measures were obtained by taking the maximum peak during a response period as a proportion of the response amplitude before the onset of the stimulus. Simple millimeter measures of the records were used in both cases.

[5] The remaining analyses, including four variance summaries and four tables of means, have been deposited with the American Documentation Institute. Order Document No. 5848, from ADI Auxiliary Publications Project, Photoduplication Service Library of Congress, Washington 25, D. C., remitting $1.25 for 35 mm. microfilm, or $1.25 for 6 by 8 in. photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

The analysis followed that suggested by Edwards (1950) for the situation where repeated observations are made on several groups of *S*s. The interaction between individual *S*s and the stimulus conditions within a given group was taken as the error term for comparison between sets of columns. Mean squares based on row comparisons are divided by a combined "between subjects" term to obtain the appropriate *F* value.

TABLE 1
MEAN LEVELS OF FINGER PULSE (FP) RESPONSE TO THE SINGLE-PERSON SLIDES

Subject Variables	Stimulus Variables					
	White Slides			Negro Slides		
	White Female	White Male	Mean for White Slides	Negro Female	Negro Male	Mean for Negro Slides
Male: Prejudiced	.755	.953	.854	.953	.978	.966
Unprejudiced	.924	1.690	1.307	1.135	.970	1.052
Female: Prejudiced	1.224	1.376	1.300	1.744	1.252	1.498
Unprejudiced	.990	1.018	1.004	.873	.928	.900
Mean for males	.840	1.322	1.081	1.044	.974	1.009
Mean for females	1.107	1.197	1.152	1.309	1.090	1.200
Mean for prejudiced	.989	1.164	1.077	1.348	1.115	1.232
Mean for unprejudiced	.957	1.354	1.115	1.004	.949	.976

Note.—Lower numerical values indicate greater physiological response (see footnote 6). Relative response ranges can be interpreted as follows: Response levels before each stimulus presentation equal 1.000. A value such as 1.744 (prejudiced females responding to Negro female slides) indicates a 74% response change in the direction of decreased FP activity.

paring the "mean for white slides" with the "mean for Negro Slides" for each of the four *S* types (Table 1). Prejudiced males and prejudiced females both gave smaller FP reactions to the Negro slides than to the white slides.[6] On the other hand, unprejudiced males and unprejudiced females showed no such decrease in level of FP response toward Negro slides, actually showing a higher level of FP activity. This comparison suggests an *inverse* relationship between degree of prejudice and autonomic response as measured by FP activity. This relationship is complicated, however, by the fact that the two sexes reacted generally with somewhat different levels of activity. That is, while the direction of response was similar for prejudiced persons of both sexes, the greatest degree of FP activity was shown by prejudiced males while viewing white female slides, and the smallest response was by prejudiced female *S*s viewing Negro female slides.

[6] Unusually high numerical values found in such cases as prejudiced females viewing Negro female slides (Table 1), need careful interpretation. The 1.744 numerical value represents a 74% decrease in FP activity over the preceding 30-sec. rest period. This seems difficult to account for in terms of relaxation of the arterioles of the finger. Rather, it would seem that the stroke volume of the heart had increased to a point where it more than cancelled the constriction response. If this is the case, the suggested inverse relationship between verbal attitude and FP activity may prove with further research to be a more complicated product of heart action and arterial constriction; a product whose components are each related to the verbal attitude variable in a different way. Independent measures of heart stroke were attempted in the present research but proved unusable. Further research on this problem should give careful consideration of this possible alternative interpretation of the suggested inverse relationship between FP activity and attitudinal predisposition.

TABLE 2
ANALYSIS OF VARIANCE FOR FINGER PULSE RECORDS, SINGLE-PERSON SLIDES

Source of Variation	*df*	Mean Square	*F*
Between subjects	33		
a. Between subject types:	3		
1. Between sexes	1	.83	.929
2. Between attitude types	1	.46	.515
3. Sex × attitude interaction	1	4.26	4.770*
b. Between subjects of same type	30		
1. Prejudiced males	7		
2. Unprejudiced males	7		
3. Prejudiced females	9		
4. Unprejudiced females	7		
Within subjects	102		
a. Between stimulus conditions	3		
1. White vs. Negro slides	1	.00	.000
2. Male vs. female slides	1	.11	.222
3. Race × sex interaction	1	1.67	3.370
b. Interactions, subjects, and stimulus conditions	99		
1. Interaction between subject, *types* and stimulus conditions	9	.36	
2. Pooled interaction between individual subjects and stimulus conditions	90	.495	
Total	135		

* Indicates significance at or beyond .05 level.

The analysis of the GSR means (Table 3) for the single-person slides reveals that the prejudiced group tended to give greater levels

TABLE 3
MEAN LEVELS OF GSR RESPONSE, SINGLE-PERSON SLIDES, LAST ⅗ OF STIMULUS

Subject Categories	Stimulus Conditions					
	White Slides			Negro Slides		
	White Female	White Male	Mean for White Slides	Negro Female	Negro Male	Mean for Negro Slides
Males: Prejudiced	.948	.981	.965	.932	.957	.940
Unprejudiced	.920	.934	.927	.840	.901	.870
Females: Prejudiced	.901	1.058	.980	.886	.953	.920
Unprejudiced	.928	1.111	1.020	.892	.879	.886
Mean for males	.934	.958	.946	.886	.929	.905
Mean for females	.915	1.085	1.000	.889	.916	.903
Mean for prejudiced	.925	1.020	.973	.909	.955	.930
Mean for unprejudiced	.924	1.022	.973	.866	.890	.878

of autonomic response to the Negro slides than did the unprejudiced group, suggesting a *direct* relationship between prejudice and GSR. In each of the three parallel analyses previously mentioned, a similar pattern of significant results appeared.

Table 3 also reveals that GSR responses by all *S* types were consistently greater to slides portraying males than to those showing females. The same situation prevails in all three analyses of the single-person slides.

In summary of the FP and the GSR analyses of the single-person slides, prejudiced *S*s manifested different levels of autonomic activity than unprejudiced *S*s while viewing Negro slides as compared to white slides. Prejudice and FP activity tended to be inversely related, while prejudice and GSR activity tended to be directly related. In addition, the sex of the *S* influenced the level and direction of response.

We turn now to a summary of the analyses of the pair slides, restricting our discussion to the GSR data.

Table 4 shows the mean levels of response toward the various types of slides. A significant *F* value (see Table 5) was obtained for GSR reactions to slides portraying persons of the "same sex" versus those presenting "mixed sex" pairs. The means of the lower part of Table 4 show that *S*s gave greater responses to slides portraying persons of the same sex.

It should be kept in mind, however, that by the time the pair slides were presented to the *S*s, they had already reacted to the set of single-person slides. The *S*s tended generally to respond with lesser magnitude, indicating a "jading" effect had set in by the time that the pair slides were presented.

The statistically significant findings of the experiment can be listed briefly as follows:

Finger Pulse Analysis

1. The degree of autonomic response accompanying exposure to the slides was related to the sex and attitudes of the *S*s taken in combination. The response of greatest magnitude was by prejudiced males viewing white female slides. The response of smallest magnitude was by prejudiced females viewing Negro female slides (Table 1 and 2).

2. Prejudiced *S*s of both sexes showed lesser levels of FP activity while viewing Negro slides as compared to white slides, while unprejudiced *S*s of both sexes showed greater reactions to Negro slides as compared to white slides, suggesting an inverse relationship between FP activity and prejudice (Tables 1 and 2).

Galvanic Skin Response Analysis

1. Prejudiced *S*s, in the single-person slide analysis, showed greater responses toward Negro slides than did unprejudiced *S*s (Table 3).

2. The *S*s as a whole, regardless of sex or attitude, gave consistently greater responses to slides portraying males than to slides portraying females (Table 3).

3. In the analysis of the pair slides, the *S*s as a whole showed higher levels of autonomic

TABLE 4

MEAN LEVELS OF GSR RESPONSE, PAIR SLIDES LAST ⅔ OF STIMULUS

Subject Variables	Stimulus Variables					
	Same-Race Slides			Mixed-Race Slides		
	Same Sex	Mixed Sex	Mean for Same Race	Same Sex	Mixed Sex	Mean for Same Sex
Males: Prejudiced	1.030	1.037	1.033	1.256	.935	1.096
Unprejudiced	1.046	.979	1.013	1.015	.949	.982
Females: Prejudiced	1.004	1.017	1.001	.962	.993	.976
Unprejudiced	.953	.946	.950	1.173	.932	1.053
Mean for males	1.038	1.008	1.023	1.136	.942	1.039
Mean for females	.979	.982	.981	1.068	.963	1.015
Mean for prejudiced	1.016	1.027	1.022	1.109	.964	1.036
Mean for unprejudiced	1.000	.963	.982	1.094	.941	1.018

	Same-sex slides	Mixed-sex slides
Mean for prejudiced	1.062	.991
Mean for unprejudiced	1.042	.952
Mean for males	1.037	.925
Mean for females	1.023	.972

TABLE 5

SUMMARY OF SOURCES OF VARIANCE FOR ALL ANALYSES

Physiological Variable	Slide Stimuli	Stimulus Period	Significant* Source of Variance	Direction of Relationship
FP	Single-person slides	First ⅓ of period		(See Tables 1 and 2)
GSR	Single-person slides	First ⅓ of period	White vs. Negro slides	Average *S*'s response greater to white slides with prejudiced *S*'s response greater to Negro slides
			Male vs. Female slides	Response greater to male slides for all *S* types
GSR	Single-person slides	Last ⅔ of period	White vs. Negro slides	Average *S*'s response greater to white slides with prejudiced *S*'s response greater to Negro slides
			Male vs. Female slides	Response greater to male slides for all *S* types
GSR	Single-person slides	Entire period	White vs. Negro slides	Average *S*'s response greater to white slides with prejudiced *S*'s response greater to Negro slides
			Male vs. Female slides	Response greater to male slides for all *S* types
GSR	Pair slides	First ⅓ of period	none	
GSR	Pair slides	Last ⅔ of period	Same-sex vs. mixed-sex slides	Response greater to same-sex slides for all *S* types
GSR	Pair slides	Entire period	none	

* Significant at the .05 level.

response toward slides portraying persons of the same sex as compared to mixed sex pairs (Table 4).

DISCUSSION

How do these findings bear on the problem as stated originally? The most simple and general answer that can be given to this question is that responses to attitude objects include nonrandom autonomic physiological patterns that are related to characteristics of the *S*, characteristics of the object, and combinations of these characteristics. There is evidence that the attitudinal predispositions of our *S*s, as measured by our attitude scales, were related to some extent to their autonomic response patterns. There is also ample evidence that the response patterns were complicated by the influence of other variables. The sex category of the *S* operated in various ways to determine partially the direction and magnitude of response. In addition, a person's autonomic responses were influenced by the racial and sexual categories of the attitude object, and the ways in which attitude objects were paired together.

What do these data mean in terms of traditional categories of affective orientation? It cannot be assumed that conspicuous physiological changes manifested by an *S* upon being exposed to a stimulus object indicates simple favorability or unfavorability toward the attitude object, regardless of his scale-measured attitudes toward that object. It is easy to fall into thinking, especially when the stimuli are "race objects," that high response indicates a negative attitude. Actually, such responses could be due to a number of conditions. Conflicting feelings, indecision, or even positive involvement might result in large autonomic responses. There are probably many other interpretative possibilities besides these. The present research makes no attempt to assign simple meaning to the responses. Our aim was to determine if measurable patterns of autonomic physiological response exist and, if so, are they related to scale measured attitudes?

Finally, in attempting to evaluate the experimental data, two aspects of our method should be kept in mind. First, it would have been easy to design the slide stimuli so that larger autonomic responses were evoked. It will be recalled that our slide stimuli were deliberately made "conservative." That is, suggestive poses, body contact, unusual clothing, or facial expressions were deliberately avoided, so that if autonomic responses occurred it came in spite of the avoidance of these secondary cues. Further explorations of this problem could well include more open portrayal of emotion-arousing interracial scenes. There is every reason to presume that response consistencies under such circumstances would be more clear-cut. Another variable that might well be introduced is the influence of well-defined social roles. For example, the Negro man and white girl portrayed together might evoke entirely different response patterns depending upon whether they were shown lying on a beach together clad in bathing suits, or talking together in a church with the Negro in a clergyman's robes. Such social roles were not included in the present research, but they may importantly influence attitudinal responses.

The second aspect of method which may have placed limitations on the magnitude of response is the order in which the slides were presented. Every *S* viewed the four single-person slides before being presented the pair slide stimuli. It is clear that there was a tendency for GSR activity to become dampened with rapidly presented stimuli. Since the data for the pair slides were collected during the last half of the laboratory session, differences between reactions might tend to be obscured. Further experimentation should use a reduced number of stimuli. One or two preliminaries at the start, followed by not more than four or five should be a maximum.

In terms of their implications for the attitude concept, the over-all results indicate that attitudinal responses include autonomic physiological activity that is related to the attitudinal characteristics of the individual as measured by more standard means.[7] They are also related to the sex of the responding individual and to features of the object of response.

[7] A recent article by Rankin and Campbell (1955) analyzes GSR responses of male students to a Negro by the ingenious device of having a Negro experimenter touch the *S*s while "adjusting" the apparatus. This problem of attributing the response to the racial aspects of the situation vs. a reaction to a particular person is discussed.

Clearly, there is little point in attempting to interpret these physiological responses as the "true attitude." Rather, it seems more useful to regard these responses as another dimension of attitudinal behavior to be considered along with the verbal and overt action dimensions.

Although one need not assume that this dimension must correspond with other dimensions in a one-to-one fashion, it would seem theoretically useful and empirically possible to establish the nature of the relationship between the various dimensions for particular populations under given circumstances. Our findings, however modest, indicate that further research along these lines may prove useful not only for clarifying the construct "attitude," but also for increasing our knowledge of the empirical referents of the construct.

SUMMARY

The problem was to determine if autonomic response was associated with exposure to objects toward which individuals have (scale-measured) attitudes. That is, do persons who are unfavorable in their verbal response to Negroes also manifest different autonomic responses to slides portraying Negroes than do those more favorable in their verbal response? The 46 subjects, half of whom were males and half females, were divided into a prejudiced and unprejudiced group on the basis of a verbal attitude test. They were then exposed to photographic slides portraying Negroes and whites in various combinations of race and sex. The finger pulse and GSR activity of each subject was recorded during these presentations.

The data indicate that greater GSR responses (but smaller FP responses) were given to Negro slides by prejudiced subjects, but the autonomic activity was influenced by the sex of the subject as well as the race and sex characteristics of the stimulus slides. Such autonomic activity may be considered as another dimension of attitudinal behavior to be considered along with the verbal and overt action dimensions. Additional research is needed before simple meaning can be assigned to the involvement of autonomic activity in attitudinal behavior.

REFERENCES

ALLPORT, G. W. Attitudes. In C. Murchison. (Ed.), *Handbook of social psychology*. Worcester, Mass.: Clark Univer. Press, 1935.

CAMPBELL, D. T. The indirect assessment of social attitudes. *Psychol. Bull.*, 1950, **47**, 15–38.

DAVIS, R. C., SIDDONS, G. F., & STOUT, G. L. Apparatus for recording autonomic states and changes. *Amer. J. Psychol.*, 1954, **67**, 343–352.

EDWARDS, A. L. *Experimental design in psychological research*. New York: Rinehart, 1950.

GREEN, B. F. Attitude measurement. In G. Lindzey (Ed.), *Handbook of social psychology*. Cambridge, Mass.: Addison-Wesley, 1954.

RANKIN, E., & CAMPBELL, D. T. Galvanic skin response to Negro and white experimenters. *J. abnorm. soc. Psychol.*, 1955, **51**, 30–33.

WESTIE, F. R. A technique for the measurement of race attitudes. *Amer. sociol. Rev.*, 1953, **18**, 73–78.

Received January 31, 1958.

EFFECT OF CHANGES IN SELF-ESTEEM UPON IMPULSIVENESS AND DELIBERATION IN MAKING JUDGMENTS[1]

JAMES E. DITTES

Yale University

SEVERAL important and frequently studied phenomena involving judgments depend for their effect upon a variable that is usually taken for granted and assumed to be constant: the occurrence of cognitive closure. All but the most trivial situations that call for judgment are characterized by ambiguity and incompleteness in the information that can be grasped initially and by the possibility of continuing assimilation of information. The state of ambiguity and the assimilation of information end when a person forms a closure, that is, when he stops deliberating and perceives a clear structure, coherence, or meaning.

Studies of conformity or rigidity generally present an individual with a piece of information and determine how closely his judgment approximates some previous judgment, either his own or another's. Variations in the readiness with which the subject forms any judgment at all, however, are generally overlooked. But without a closure, the apparent rigidity or conformity could not occur, and there exists the strong possibility that the two variables are frequently confounded. A finding of strong conformity may mean no more than that the person is motivated to form a closure impulsively and that he imposes on the material the most easily accessible structure or meaning, namely that provided by another.

Similarly, other phenomena, such as field dependence, suggestibility, ethnocentrism, resistance to free association in psychotherapy, several types of intellectual inefficiency, and some characteristics of obsessive behavior may frequently be confounded with the variable of closure and may be related as members of the "closure family." The possibility appears great enough to warrant an attempt to study the closure phenomenon itself and to discover some of its determinants, preliminary to an attempt to unravel it from "pure" conformity, "pure" rigidity, etc.

Several lines of research have called attention to impulsiveness of closure and have attributed it to "effort after meaning" (Bartlett, 1932) or "intolerance of ambiguity" (Cohen, Stotland, & Wolfe, 1955; Frenkel-Brunswik, 1949; Smock, 1955). Such a conceptualization tends to isolate one extreme of a variable and to regard it as a discrete characteristic, fairly stable within any individual and diagnostic of particular personality characteristics. Closure may be impulsive or delayed, but in every judgment it occurs. It is proposed here to regard impulsiveness or delay as representing differing strengths of a general response tendency to achieve closure. These strengths can be measured by differing latencies in situations that also arouse motives that oppose the closure tendency—e.g., the desire for a qualitatively superior judgment. A strong response tendency produces closure at a lower threshold—i.e., with less information assimilated—than does a weaker response tendency.

Two determinants of the strength of closure are considered in this study. One is based on the assumption that the act of closure commonly acquires a reward value as a means of enhancing self-esteem. Self-esteem is regarded as a general sense of self-assurance or of adequacy, depending on a variety of internal and external stimuli (including the approval of others, achievements, and reassuring self-verbalizations), all of which are designated here as sources of self-esteem. Any damage to self-esteem, it may be hypothesized, increases the strength of the tendency to achieve closure as one means of restoring or enhancing self-esteem. The other determinant is a personality characteristic designated as general desire for clarity.

The assumptions and hypotheses guiding this research may now be stated more explicitly.

Assumption 1. The relation between self-

[1] Based on a doctoral dissertation at Yale University. Much appreciation is due members of the committee, Arthur R. Cohen, George F. Mahl, and Leonard W. Doob, Chairman. Robert Hyatt rendered valuable assistance in conducting the experimental sessions.

esteem and its sources is a negatively accelerating increasing function, approaching asymptote (see Fig. 1). If self-esteem is low, a single source provides a relatively great increment. If self-esteem is already high, the same source provides a relatively small increment.

Assumption 2. The desire to increase self-esteem functions as a drive, the strength of which may be represented by the difference between the existing amount of self-esteem and the projected maximum. Various strengths of such a drive are represented in Fig. 1 as vertical distances between the curve and the level of asymptote. This desire to increase self-esteem motivates the seeking and acquisition of additional sources. Which particular sources a person seeks depends on (*a*) the type of sources that have acquired self-esteem value for him and which he has been reinforced in seeking and (*b*) opportunity afforded by existing circumstances.

Assumption 3. The act of achieving closure is one commonly acquired source of self-esteem. Persons feel more self-assurance when they have a cognitive mastery of a situation. This reward value of closure is the result of previous experience in a wide variety of circumstances in which important need satisfactions are dependent upon cognitive closure. Such experiences may be as varied as the child's learning to name objects for the sake of parents' approval to the reduction of tension that follows the labeling or clarification of emotions in psychotherapy. They include all the instances in which action leading to the gratification of needs, even as basic as hunger, is dependent on a previous meaningful appraisal of the stimulus field. Individual and cultural differences in the amount of reward value that closure carries are to be expected, depending upon differences in learning experiences due to such factors as variation in parents' tendency to reward cognitive achievements. But the wide adaptive value of closure probably results in the acquisition of a generally high reward value in any culture.

This theoretical analysis makes the simpler assumption that closure is a source and that it provides an increment of self-esteem. The actual facts are probably more complicated. The presence of ambiguous, unstructured

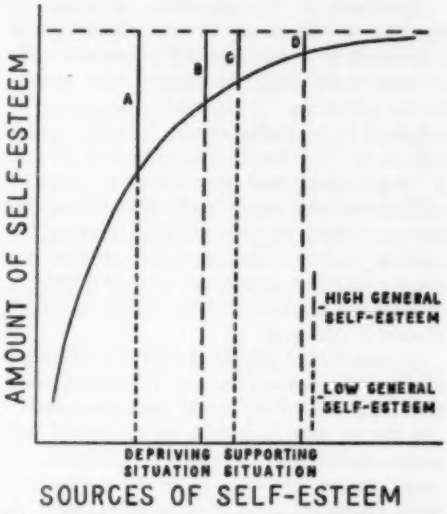


FIG. 1. REPRESENTATION OF ASSUMED RELATION BETWEEN SOURCES AND AMOUNT OF SELF-ESTEEM

(Assumed strength of motivation to increase self-esteem is represented by the distance between the curve and the level of asymptote. Representative strengths are indicated by lettered lines for persons with high and low levels of initial self-esteem in depriving and supporting situations.)

stimuli undoubtedly acquires threatening or anxiety-arousing properties and, in effect, produces a decrement in self-esteem. The derivation of the hypotheses would not be substantially affected if this assumption were made instead of, or in addition to, the assumption that closure increases self-esteem.

Assumption 4. Among other determinants of the strength of the closure tendency, there seems to be a set of persistent and general behavioral tendencies that may be designated as a *desire for clarity*. Such a trait may be indicated by the desire for explicit goals and purposes, for routine in daily tasks, for precision of thought, and for unambiguity in interpersonal relations. The desire for clarity may be regarded as a general "life-style" with diverse origins and functions, not considered here, and firmly enough entrenched to operate relatively autonomously of any situational variation. The strength of the closure tendency is assumed to be a direct function of the strength of the desire for clarity.

Hypothesis I: Persons whose self-esteem is damaged or threatened by the withdrawal or deprivation of a source of self-esteem are likely to show more impulsive closure than persons whose self-esteem is relatively guaranteed or enhanced by a similar source. This hypothesis follows directly from Assumptions 1, 2, and 3. Deprivation increases drive to enhance self-esteem and hence increases striving for the acquisition of any available sources, including closure. The increased striving for closure, in effect, establishes a lower threshold, in terms of amount of information, at which closure is achieved.

Hypothesis II: Among persons experiencing a situationally induced loss of self-esteem, those with lower general self-esteem show more impulsive closure than those with higher general self-esteem. Aside from any single situational influence, persons probably maintain a relatively even, persistent level of self-esteem, which is related to impulsiveness of closure independently of situational differences. This relation is more likely to be apparent among persons in a situationally deprived condition than among those whose self-esteem is enhanced in the situation. When the situation is supporting, all self-esteem is pushed more or less uniformly high—on the flat part of the curve in Fig. 1. When the situation is depriving, however, the more persistent personality differences are effective. This relation is illustrated by the greater difference between Lines *a* and *b* than between Lines *c* and *d* in Fig. 1.

Hypothesis III: The difference predicted by Hypothesis I is greater among persons with generally low self-esteem than among persons with generally high self-esteem. If general self-esteem is high, the addition or withdrawal of a single source within the situation has relatively little effect. If it is low, the variation of a single source has greater affect. In Fig. 1, the difference between Lines *a* and *c* is greater than the difference between Lines *b* and *d*.

Hypothesis IV: Among persons not experiencing a situational deprivation of self-esteem, those with greater general desire for clarity show more impulsive closure than those with less desire for clarity. This prediction follows directly from Assumption 4. Even in the absence of situational conditions motivating impulsiveness of closure, a strong general desire for clarity may provide the motivation. However, when there is a strong motivational arousal of impulsiveness, as hypothesized for the condition of deprived self-esteem, the correlation between impulsiveness of closure and desire for clarity is likely to be obscured.

Hypothesis V: The difference predicted by Hypothesis I is greater among persons with lower general desire for clarity than among persons with high desire for clarity. If Hypothesis IV holds, it necessarily follows that the situational difference tends to be effective only among persons with low desire for clarity.

The last four hypotheses, concerning the interaction between the situational conditions and the two personality characteristics, self-esteem and desire for clarity, may be summarized as follows: When either personality characteristic is weak, the situation difference has more effect, since then a person is more vulnerable to situational variation (Hypotheses III and V). When the situation supplies strong stimulation that has the same effect as the personality characteristic—i.e., when it supports self-esteem or when it provides incentive to closure—the effects of differences in the personality characteristic are obscured. When the situation is negative with respect to the personality characteristic—i.e., when there is deprivation of self-esteem sources or when there is an absence of incentive to closure—the personality difference has greater effect (Hypotheses II and IV).

METHOD

Subjects

One hundred and four volunteers from the Yale freshman class met in 20 small groups of up to six members each for what they believed was a session intended to demonstrate good discussion techniques. Experimental procedures were replicated in each group.

Ss' motivation was heightened by offering a prize of $60 and prestigeful recognition to the discussion group performing the best, by letting the *Ss* believe that the discussion and task materials assigned their group were chosen to meet topic preferences that they had expressed during recruiting, and by stressing the importance of the research. A tape recording was made of each group's discussion to emphasize the impression that their discussion would be studied later by social scientists.

Manipulation of Self-Esteem

The first half of each session was devoted to a discussion of juvenile delinquency, based on case material

presented to the group. This discussion was interrupted three times to allow members to rate each other anonymously on the following scale:

How desirable is it that this person be kept in the group?
__extremely desirable
__very desirable
__somewhat desirable
__not very desirable, but he should be kept in
__he should be rejected from the group.

As a rationale for the request for ratings, it was explained that the investigator was interested in knowing how individual members appeared to adjust to such a situation, and that such ratings would also enable the group to improve its performance by eliminating any members who appeared detrimental to its success. *E* thumbed through the ratings as they were made and discarded them in a wastebasket.

During an intermission after the discussion of delinquency, *E* turned the conversation to these ratings and, after eliciting *S*s' interest in seeing them, he retrieved the slips from the wastebasket and distributed them. Each *S* was thus allowed to see privately what he thought were the ratings made of him by the other members. These were, in fact, fictitious ratings prepared in advance and substituted for the originals in the wastebasket. No *S* ever indicated doubt of the genuineness of the ratings.

So that *S*s would similarly interpret the ratings, *E* announced that he had noticed the average ratings to be around "very desirable." In each group, some *S*s found their ratings clustered at "extremely desirable," some at "very desirable," and some at "somewhat desirable." The conditions were assigned at random around the discussion table and varied for each session.

The three conditions were intended to induce feelings of high, average, and low acceptance. Evidence of the *S*s' direct reactions to the conditions was obtained on a questionnaire at the conclusion of the session and is reported in Table 1. *S*s in the high and average conditions did not differ substantially in their reactions. They both reported being pleased rather than disturbed by the ratings, and on the other measures they showed similar reactions. Consequently, the data for these conditions are pooled in reporting results and regarded as a single condition of high acceptance. As compared with the high and average conditions, *S*s in the low condition felt disturbed, less accepted, and less adequate; and they engaged in the defensive maneuver of devaluing their membership in the group.

A second type of threat to self-esteem was also introduced. At each of the levels of acceptance, half the *S*s were given the impression that the evaluation by their peers was stable and unlikely to change, and half believed that the evaluation was conditional and very likely to change. An unstable rating was expected to threaten self-esteem and produce the same closure behavior as predicted for low acceptance. Direct measures indicated that this manipulation was less clearly disturbing to *S*s, and it produced less clear differences in closure. Results among *S*s who felt threatened by unstable ratings generally achieved statistical significance in the predicted direction. These are reported in Dittes (1958). In this report, the stable and unstable conditions are simply combined at each level.

TABLE 1
DIRECT EFFECTS OF EXPERIMENTAL CONDITIONS

Measure[a]	Experimental Conditions			p Value
	High	Average	Low	High & Average vs. Low
Recall of received ratings	3.5	2.6	1.8	$<.01$
Pleased or disturbed	4.0	2.0	−.8	$<.001$
Sense of adequacy among peers	4.5	4.7	3.8	$<.02$
Valuation of group membership	5.0	5.2	4.3	$<.005$

[a] Typical questionnaire items for each measure:
Recall: "How do you think the other members of the group rated you?"
Pleased or disturbed: "Were you pleased or disturbed when you saw the ratings?"
Sense of adequacy: "In other groups you've been in, how much do others generally like working with you and having you in the group?"
Valuation of membership: "If this group should be invited to meet again or to serve as a demonstration panel, would you want to continue working with it?"

Measures of Impulsive Closure

Parable Task. *S*s were presented with a one-paragraph story, written in Biblical idiom and represented as a portion of a recently discovered scroll, and asked to write their opinions concerning the meaning of the parable. The account was essentially incoherent and unstructured. However, it contained a large number of familiar religious symbols that could, with some selection and distortion, be forced to yield a coherent meaning, though no two *S*s ever found the same meaning. Subsequently, two judges decided, on a three-point scale, whether the response was a positive statement of meaning or an indication of uncertainty as to meaning. As for all the tasks, the scoring categories were defined in advance of the experimental sessions, and the judges worked independently, without knowledge of the *S*s' performance on other tasks or of the experimental conditions to which the *S*s had been assigned. The categories were so defined that the higher score represented more impulsive closure. The judges showed absolute agreement on this task in 86% of the judgments. The contingency coefficient representing the agreement in sorting into the three categories was .75.

Impression Formation Task. *S*s were given two lists of traits that were represented as words used by close friends to describe two students; they were instructed to write their impression of each man. Each list contained a dominant set of traits and a subordinate set which was incompatible with the dominant set. In one instance, for example, the list was: individualist, suspicious, belligerent, considerate, self-concerned, thoughtful, stubborn, tolerant. The two lists were balanced so that dominant traits in one were favorable, in the other

unfavorable. Subsequently, two judges decided, on a three-point scale, the extent to which the subordinate traits were integrated into the impression. A statement based exclusively on the dominant traits was regarded as the most impulsive. There was absolute agreement on 72% of the judgments, and the agreement yielded a contingency coefficient of .64.

Jammed Message Task. *Ss* listened to a sound tape on which a phrase was repeated 20 times with decreasing volume of masking noise on each repetition. They indicated the number of the repetition on which they were "quite sure" that they recognized the phrase by selecting the correct phrase from among five alternatives read before the tape was played. The complete task comprised three such phrases. The measure of impulsiveness was the repetition on which the *S* first claimed to recognize the phrase, this summed for the three phrases. Because of significant differences between groups, the deviation of each *S* from the mean of his group was taken as his score, with a positive sign given to the earlier response.

Summary of Task Characteristics. Each task presented the *S* with unstructured, ambiguous elements that he regarded as potentially coherent and meaningful. Each task presented cues that provided the opportunity for an immediate structuring; for the three tasks these were, respectively, commonly used religious symbols, a quickly apprehended set of dominant traits, and fragments of a spoken phrase. Each task also provided cues prompting delay and withholding of early closure. These were of two types: (*a*) information that was not readily assimilable and therefore had to be ignored or distorted if a quick closure was to be achieved; and (*b*) indications that the available information was incomplete and that more was forthcoming. The first type was represented in the parable and impression formation tasks by the presence of elements incompatible with the kind of judgment that could be formed impulsively. The second type was represented in the parable by the announcement that it was but a portion of a translation and in the jammed message task by the knowledge that the phrase on the tape was about to be heard more clearly in another 15 seconds.

Each task determined individual differences in threshold of closure. The parable task presented a constant stimulus situation, approximately at the average threshold and determined for each *S* whether he formed a closure. The other two tasks presented gradually assimilable information and determined how much information each *S* assimilated before he formed a closure.

Combined Index. Since all the tasks were generated from the same rationale to measure a single behavioral tendency, the scores for each task were combined to form a more reliable over-all index of closure strength for each *S* by summing the scores after weighting them by factors to equalize their variance. The low though consistently positive correlation among the task scores (ranging from .06 to .19) is presumably due to the fact that each task elicits peculiar skills, interests, or background that override the experimental effects for individual *Ss*, but still leave important group differences.

Personality Measures

General Desire for Clarity. Six items on a questionnaire before the experimental session sampled the *Ss*' desire for clarity in a variety of life situations. Some items were taken from a scale used by O'Connor (1952) and some from a questionnaire used by Cohen et al. (1955). The items were as follows:

1. It is always better to have a definite course of action rather than to be vacillating among several possibilities.
2. A smart person gets his life into a routine so that he is not always being bothered by petty details.
3. In writing a theme or essay, clear organization of the material is of utmost importance.
4. A committee works best when the responsibilities of each member are clear.
5. It is better to keep on with the present method of doing things than to take a way that might lead to chaos.
6. Textbooks which are unclear are really useless, no matter how complete or otherwise valuable they may seem to be.

Ss indicated agreement or disagreement with each item on a six-point scale. Reliability is indicated by a product-moment correlation, corrected for length, of .84 between the scores of the odd- and even-numbered items.

For purposes of comparison, the distribution of scores was divided into quarters at the most natural cut-off points, before any of the results were examined. No alternative division was undertaken. The self-esteem scores were divided into thirds on the same basis; since these scores tended to be tri-modal, to have separated them into quarters would have arbitrarily divided clusters of scores.

Self-esteem. General self-esteem was measured with an index made up of three components. (*a*) Eleven self-rating items were used on the questionnaire before the experiment. Most of these were adapted from Murray (1938) as was the form of administration. The *S* indicated on a six-point scale whether each of the items applied more or less to him than to the average man his age. Typical items were:

TABLE 2
AVERAGE CLOSURE SCORES

Task	Experimental Conditions			*p* Value
	High	Average	Low	High & Average vs. Low
Parable	2.0	2.0	2.5	<.05[a]
Impression formation	2.1	2.2	2.3	<.35
Jammed message	−1.2	−.7	1.9	<.05
Combined index	43.6	46.5	54.5	<.002

[a] Based on a chi-square test of the distribution among the three scoring categories, because scores were not normally distributed. Other tests are *t* tests.

I feel embarrassed and uncomfortable in the presence of people who are socially gifted.

I feel anxious and uncertain when I am suddenly faced by a critical situation.

The product-moment correlation between the scores on the odd- and even-numbered items, corrected for length, was .82.

(*b*) A measure of sense of adequacy among peers was obtained with four items on the questionnaire at the end of the session. This score was found to be significantly correlated ($r = .34$, $p < .01$) with the before-session measure of self-esteem, after differences attributable to experimental manipulation were removed by taking as each *S*'s score his deviation from the mean of his experimental condition.

(*c*) The actual ratings that *S*s received from each other, though never actually saw, were justified as another indication of level of self-esteem by the assumption that because of the brief period of acquaintance preceding the actual ratings, these were likely to be based on behavior (e.g., active participation vs. reticence) which is largely a function of a person's own appraisal of himself. This measure correlated significantly ($r = .25$, $p = .01$) with self-ratings.

These three measures had almost identical variances and were summed to provide the combined index. The decision to use this combined index to measure self-esteem was made before the results were examined.

RESULTS

Hypothesis I. Persons made to feel well accepted consistently showed less impulsiveness than those made to feel poorly accepted. Table 2 reports the results for each of the three tasks and for the combined index, which showed a highly significant difference ($t = 3.32$, $p < .002$)[2] between the high and average and the low conditions. The intended high and average conditions, which showed similar direct reactions to the experimental manipulation as reported above, also produced substantially similar results, though tending consistently to differ in the predicted direction.

Results relating to the hypothesized interaction of the experimental conditions with the personality characteristics are reported for the combined index of closure in Tables 3 and 4.

Hypothesis II is tentatively supported by the second column in Table 3. Within the condition of low acceptance, the third of the *S*s with lowest self-esteem were more impulsive

[2] Throughout, all probability values are derived from using both tails of the statistical distribution.

TABLE 3

AVERAGE SCORES ON COMBINED INDEX OF CLOSURE AT DIFFERENT LEVELS OF SELF-ESTEEM

Level of Self-Esteem	High Acceptance	Low Acceptance	Difference	p Value
High	47.3	52.6	5.3	$<.30$
Average	42.3	52.1	9.8	$<.05$
Low	45.8	58.8	13	$<.05$

TABLE 4

AVERAGE SCORES ON COMBINED INDEX OF CLOSURE AT DIFFERENT LEVELS OF DESIRE FOR CLARITY

Level of Desire for Clarity	High Acceptance	Low Acceptance	Difference	p Value
High	49.2	53.8	4.6	$<.40$
High average	44.2	51.9	7.7	$<.20$
Low average	44.4	53.8	9.4	$<.10$
Low	43.5	58.0	14.5	$<.01$

TABLE 5

AVERAGE CLOSURE SCORES ON SEPARATE TASKS FOR DIFFERENT LEVELS OF SELF-ESTEEM AND DESIRE FOR CLARITY

	Parable Task				Impression Formation Task				Jammed Message Task			
	High Acc.	Low Acc.	Diff.	p	High Acc.	Low Acc.	Diff.	p	High Acc.	Low Acc.	Diff.	p
Self-esteem:												
High	1.1	1.3	.2		2.2	2.2	0		− .5	3.6	4.1	
Average	.8	1.6	.8	$<.05$	2.1	2.2	.1		−1.5	.2	1.7	
Low	.9	1.7	.8	.10	2.2	2.5	.3	.10	− .4	2.0	2.4	
Desire for Clarity:												
High	1.3	1.5	.2		2.2	2.4	.2		.4	− .1	− .5	
High average	.8	1.5	.7		2.4	2.4	0		−3.2	−1.6	1.6	
Low average	.8	1.6	.8	$<.05$	2.1	2.0	− .1		1.8	4.1	2.3	
Low	1.0	1.6	.6		2.0	2.2	.2		−2.4	4.8	7.2	.02

($t = 1.7$, $p < .10$) than the upper two-thirds. Among *S*s in the high condition, no variation approached significance. *Hypothesis III* is supported by the trend of a progressively greater difference between experimental conditions in lower thirds of self-esteem.

As predicted by *Hypothesis IV*, persons with greatest desire for clarity tended to be impulsive even when they were in the experimental condition of high acceptance. Within the high condition, *S*s in the top quarter of desire for clarity were more impulsive ($t = 2.08$, $p < .05$) than those in the lower three-quarters. *Hypothesis V* is supported by the tendency for the difference between experimental conditions to become progressively greater with less desire for clarity.

Results with the separate tasks, reported in Table 5, were substantially like those with the combined index, although significance levels are lower as a result of the greater unreliability of any single estimate of closure.

DISCUSSION

Low evaluation by peers clearly results in greater impulsiveness of closure on various types of conceptual tasks. This effect depends in an important way, however, upon certain entrenched personality characteristics. The findings can be accounted for in terms of the damage to self-esteem produced by low acceptance and of the capacity to enhance self-esteem that closure commonly acquires. Impulsive closure becomes a compensating source of self-esteem.

There are various alternative accounts that can be advanced, although none of them seems congruent with all the available findings. One important possibility is that poorly accepted persons produced impulsive closure in order to regain acceptance or to forestall rejection, that is, for its direct *instrumental* value. The mediation of changes in self-esteem, it may be argued, need not be invoked. Such an explanation, however, requires certain assumptions and predictions that are not supported by the available data:

1. It must be assumed that *S*s experienced some direct association between impulsiveness of closure and acceptance by these peers. If *S*s were trying to regain acceptance by the group through impulsive closure, they must have supposed that the group (or *E*) would reward impulsiveness and punish deliberation. However, because this possibility raises an important problem in interpreting some previous similar findings (e.g., Smock, 1955), care was taken in designing the procedures to insure that impulsiveness could not become an implicit norm within these groups any more than delayed, more cautious closure. The only instructions were general suggestions to *S*s to engage in good discussion and to make good decisions, and they were left to interpret "good" in any manner they chose. The last two of the tasks were presented as incidental exercises, completely unrelated to even this group goal of good discussion, and the *S*s were clearly instructed that their performance was not to be known by the other members.

On the contrary, it seems more likely that the groups developed a norm for deliberative judgments that tended to work against the actually obtained results. The academic environment from which the *S*s were drawn and in which the experiment was held tends to reward quality of closure rather than impulsiveness on complex judgments of the sort required here. Observation of the discussion and some questionnaire responses indicate that within the experimental sessions *S*s did feel pressure toward delayed closure. These are reported elsewhere (Dittes, 1958). None of the evidence suggests that *S*s consciously experienced any pressures toward impulsiveness.

Although time pressures may have been experienced occasionally by *S*s, impulsiveness of closure is not necessarily related to response time; rather, it refers to whether the thought process occurs in an essentially structured or ambiguous framework. A person might spend much time and effort interpreting a parable, for example, by compiling meanings of the different symbols and perhaps even by shifting from one interpretation to another. But the impulsiveness exists in the primary effort of grasping for some certain meaning. Neither of two possible indications of response time, the rank orders within each group of beginning and of completing a response to the parable, was correlated with scores of impulsiveness on that task.

2. Since it is well established (Back, 1951; Dittes & Kelley, 1956; Festinger, 1950) that

efforts to regain group acceptance, such as this explanation supposes, are in part a function of *S*s' evaluation of their membership in the group, this alternative interpretation must predict that the effects would be greater among *S*s who most highly valued their membership than among those who least valued their membership. However, when *S*s are divided according to their indication on the final questionnaire of high or low evaluation of their membership in the group, virtually identical results are obtained with those who set a high value on membership as with those who set a low one. The respective scores of these two groups on the combined index under conditions of high and low acceptance were 45 and 55, and 46 and 55. The valuation measure must be regarded as reliable because, in other results obtained in the experiment and reported elsewhere (Dittes, 1958), it was shown to be useful in predicting results that could be clearly attributed to a group conformity effect.

The above discussion concerns the possible interpretation that *S*s performed impulsive closure because they believed that *impulsiveness* would be rewarded by greater acceptance. Another possibility is that *closure* has reward value, not as a source of self-esteem as posited by Assumption 3, but exclusively by association with the approval of other persons. A striving to improve acceptance, rather than a striving to improve self-esteem, may mediate the results; but if this is so, it must be regarded, for the reasons indicated just above, as functioning exclusively on the basis of learned associations and independently of any realistic expectations about potential approval or disapproval within the experimental situation.

Such a striving to improve acceptance would be functionally equivalent with striving to improve self-esteem, and if self-esteem were considered as based exclusively on the approval of others, the two assumed motives would be identical conceptually. The self-esteem hypothesis proposes that the self-esteem value of closure is based on other need gratifications as well as on praise by others and therefore that the results of this experiment would be substantially the same if self-esteem were manipulated by varying some other source, such as achievement and failure experiences.

A third possible explanation would attribute the effects to the distraction caused by the treatment of low acceptance. It is possible that poor acceptance elicits aggressive, depressive, or other reactions within *S*s which could have the effect of distracting them from full attention or motivation on the subsequent tasks. Lessened attention on the tasks, it might further be argued, would lead to off-hand, less careful responses, which would be scored as impulsiveness. Asch (1952), for example, has proposed that when attention is directed toward ego-concerns (as the condition of low acceptance may do) it is correspondingly withdrawn from the task, and this explanation has been used by Shipley (1956) to account for findings similar to those reported here. A related theory of "interference" is used by Moffitt and Stagner (1956) to account for similar findings.

There is no direct evidence available on the extent to which *S*s were so distracted. But related data suggest that distraction effects were not critically involved: In the first place, any such distraction is likely to be reflected in responses to the items measuring *S*s' motivation to be in the experimental group. One of the items, for example, asked how hard *S*s tried to perform well. Yet, as already reported, results are virtually identical among *S*s who scored low on the measure of motivation and those who scored high. Secondly, examination of the definitions of the scoring categories for the tasks does not suggest that behavior scored as impulsive is likely to be the result of casual, off-hand response nor of inattentive, inefficient, or primitive intellectual functioning, except possibly on the impression formation task. Early attempts to decipher the phrase on the jammed message task and the intellectual agility required to discover some meaning in the parable suggest greater involvement and attention, not less.

SUMMARY

Closure is defined as the structuring of otherwise ambiguous stimuli so as to provide meaning and to end deliberation. Assuming that (*a*) such closure commonly acquires reward value as a source of self-esteem through learning experiences in which it has been associated with achievements, praise by other persons, and other fundamental sources of self-esteem;

and (*b*) damage or threat to self-esteem produces a stronger motivation to increase self-esteem by various means, including closure; it was hypothesized that a situationally induced threat or injury to self-esteem would result in greater impulsiveness of closure in making complex judgments. Such impulsive closure may be the equivalent of behavior sometimes attributed to "effort after meaning" or "intolerance of ambiguity," and may be an important component of rigidity, conformity, suggestibility, field dependence, ethnocentrism and similar members of a possible "closure family."

Self-esteem was manipulated by communicating to *S*s, meeting in small groups, fictitious information as to how well each one was regarded by the others in the group. Impulsiveness or delay in closure was determined by three different behavioral measures. The threat to self-esteem of being poorly accepted resulted in significantly more impulsive closure as represented by tendencies (*a*) to find a positive meaning in an essentially incoherent prose passage rather than to acknowledge its incoherence, (*b*) to base an impression of another person exclusively on more prominent traits and to ignore inconsistent traits, and (*c*) to claim to recognize a disguised phrase, which was repeated on a sound tape with ever-increasing clarity, earlier in the sequence of of repetitions.

Interactions were predicted between the situational conditions and personality characteristics of general self-esteem and general desire for clarity, as measured by a questionnaire. Support or threat tended to be more effective in producing differences in impulsiveness of closure among *S*s with low self-esteem and among *S*s with low general desire for clarity; their behavior can be regarded as less determined by persistent personality characteristics and more vulnerable to situational differences. Personality characteristics alone tended to affect impulsiveness of closure in instances in which the situational stimulation did not obscure the personality differences: among *S*s who were situationally threatened, those with low self-esteem tended to be more impulsive than those with high self-esteem, and among *S*s situationally supported, there was an indication of a positive correlation between general desire for clarity and impulsive closure.

REFERENCES

ASCH, S. E. *Social psychology*. New York: Prentice-Hall, 1952.

BACK, K. W. Influence through social communication. *J. abnorm. soc. Psychol.*, 1951, **46**, 9–23.

BARTLETT, F. C. *Remembering*. Cambridge: Cambridge Univer. Press, 1932.

COHEN, A. R., STOTLAND, E., & WOLFE, D. M. An experimental investigation of need for cognition. *J. abnorm. soc. Psychol.*, 1955, **51**, 291–294.

DAVIDS, A. Some personality and intellectual correlates of intolerance of ambiguity. *J. abnorm. soc. Psychol.*, 1955, **51**, 415–420.

DITTES, J. E. Effects of changes in self-esteem upon impulsiveness and deliberation in making judgments. Unpublished doctoral dissertation, Yale Univer., 1958.

DITTES, J. E., & KELLEY, H. H. Effects of different conditions of acceptance upon conformity to group norms. *J. abnorm. soc. Psychol.*, 1956, **53**, 100–107.

FESTINGER, L. Informal social communication. *Psychol. Rev.*, 1950, **57**, 271–282.

FRENKEL-BRUNSWIK, ELSE. Intolerance of ambiguity as an emotional and perceptual personality variable. *J. Pers.*, 1949, **18**, 108–143.

MOFFITT, J. W., & STAGNER, R. Perceptual rigidity and closure as functions of anxiety. *J. abnorm. soc. Psychol.*, 1956, **52**, 354–357.

MURRAY, H. A. *Explorations in personality*. New York: Oxford Univer. Press, 1938.

O'CONNOR, PATRICIA. Ethnocentrism, "intolerance of ambiguity," and abstract reasoning ability. *J. abnorm. soc. Psychol.*, 1952, **47**, 526–530.

SHIPLEY, T. E. Threat to the self, the direction and breadth of attention, and the distance gradient. Unpublished doctoral dissertation, Univer. of Pennsylvania, 1956.

SMOCK, C. D. The influence of psychological stress on the "intolerance of ambiguity." *J. abnorm. soc. Psychol.*, 1955, **50**, 177–182.

Received February 19, 1958.

STABILITY AND VALIDITY OF ACHIEVEMENT FANTASY[1]

JEROME KAGAN AND HOWARD A. MOSS

Fels Research Institute at Antioch College

ALTHOUGH projective and fantasy instruments have been widely used both in clinical and research settings the reliability and validity of many fantasy variables still remains in question. This paper is the first in a series of reports dealing with the stability and validation of fantasy variables from the TAT and Rorschach and deals specifically with achievement fantasy on the TAT. Interest in achievement fantasy has been fairly intense in recent years and many studies have attempted to predict various types of achievement behaviors from fantasy responses (Atkinson: 1954, 1957; French: 1955, 1958; French & Thomas, 1958; Lazarus & Baker, 1957; McClelland & Liberman, 1949; McClelland, Atkinson, Clark, & Lowell, 1953; Vogel, Lazarus, & Baker, 1957; Winterbottom, 1952). Most workers conceive of achievement themes as a measure of the intensity of achievement motivation, and this measure has been labeled "need Achievement." However, Farber (1954) has pointed out that in an explanatory system that employs motive and response strength as independent parameters, as in behavioral theory, it may be fallacious to use fantasy behavior as a measure of motive strength. Since a fantasy response is a sample of overt behavior, it is influenced by all conditions that affect behavior, including both goal-oriented and goal-avoidant response tendencies. That is, when avoidant or inhibitory responses are associated with a motive the probability of occurrence of any motive-related response is, in part, a function of the strength of the avoidant tendencies. Thus, when one is dealing with a complex behavior like achievement, variables such as anxiety over ability to obtain an achievement goal, independent of the motivation for the goal, may elicit avoidant responses that inhibit both direct achievement behavior and achievement fantasy. On these premises, no single class of responses such as fantasy themes is adequate for assessing the intensity of achievement motivation, and it seems conceptually less ambiguous to view achievement themes as a response variable, or as one index of a person's behavioral tendency to engage in achievement behaviors. This point of view seems similar to that of Atkinson (1954) who defines "need Achievement" in response terms: ". . . the strength of a family of perceptual and instrumental response dispositions corresponding to events in past behavioral sequences which have led to the anticipated goal state . . . ." Further discussion of the theoretical interpretation of achievement fantasy will be reserved for the latter section of the paper.

This research deals with three aspects of a fantasy measure of achievement striving: (*a*) its stability over time, (*b*) its relation to certain antecedent conditions, and (*c*) its prediction of overt achievement behavior.

The Fels Research Institute is well suited to undertake a study of this kind, since for a large sample of children it has continuous longitudinal data, including TAT protocols, ratings, and reports of the mother's attitudes and behavior, school grades, and continuous IQ testing. Some of the information is relevant for the above three questions. Since the TAT cards were administered every 2½ to 3 years, the long-term stability of this behavior can be assessed. Lowell (1950), using equivalent sets of fantasy stimuli and a test-retest interval of one week, found a significant correlation between two achievement fantasy scores (phi = +.44). This report offers the opportunity to assess the stability of achievement fantasy over a much longer period of time. With respect to antecedent variables, it was felt that both direct, maternal concern with the child's achievement as well as the educational level of the parents might relate to the degree of achievement striving in the older child. It was assumed that if a parent valued achieve-

[1] This research was supported in part by a research grant (PHS-M-1260) from the National Institute of Mental Health of the National Institutes of Health, United States Public Health Service. The authors would like to acknowledge the fact that the ideas of L. W. Sontag concerning the relation between changes in IQ and personality variables was an important stimulus for the current study. We also wish to thank Vaughn J. Crandall for his careful reading of the manuscript.

ment behavior he would encourage and praise this behavior in the child and create reinforcement conditions that would facilitate the learning of the behavioral tendency. Thus, we anticipated a positive relation between ratings of early maternal encouragement and concern with a child's cognitive and motor development and later achievement strivings in the child. In addition, amount of formal education in the parent was selected as an indirect index of achievement concern on the part of the parent, since the acquiring of a high school diploma or college degree is generally viewed as an achievement-related goal. It is a relatively crude index, since many circumstantial factors are involved in determining educational level.

With respect to the behavioral consequences of achievement strivings, it was assumed that the desire to perform well in an intellectual test situation was one reflection of achievement strivings. In a recent monograph, Sontag, Baker, and Nelson (1958) reported that children who were competitive and initiated activities on their own showed large increases in IQ score during the childhood years. Kagan, Sontag, Nelson, and Baker (1958), using the same sample of children, found that those children who showed large childhood increases in IQ were more likely to tell TAT achievement themes than the children with decreases in IQ. It was suggested that the achievement fantasy reflected a tendency to strive for efficient performance in the IQ test situation and that this striving facilitated IQ gains. The present research also used IQ gain during childhood as an index of achievement strivings and offered the opportunity to study the relation between achievement fantasy and IQ changes where the changes were not extreme but representative of the entire distribution.

METHOD

From the files of the Institute population, 86 cases (*Ss*) were selected (44 boys and 42 girls) on whom the following information was available: (*a*) three seven-card TAT protocols with median ages of administration at 8-9, 11-6, and 14-6; (*b*) annual Stanford-Binet IQ scores from age 3 through 11; (*c*) educational background of each parent; and (*d*) written reports of the mother's behavior with the child based on 6 to 10 visits to the home for the first three years of the child's life. This latter information was available on only 44 cases.

Scoring of Achievement Fantasy

Each of the *S*'s three protocols consisted of seven TAT cards (Cards 1, 5, 14, 17 BM, 3 BM, 6 BM, and 3 GF), and achievement fantasy was scored according to the scheme described by McClelland et al. (1953). Since incidence of the subcategories of the McClelland scoring system were infrequent (e.g., I+, I−, Ga+, etc.), only stories in which achievement concern or behavior was the major aspect of the plot are considered. These are scored *Ach Th* in the McClelland scheme. In addition, because there was a lack of comparability among the examiners with respect to the inquiry questions, only the spontaneous verbalization of the *S* was scored. Agreement between two independent raters scoring the themes was 95%.

Educational Background of the Parents

The formal educational training of each parent was assigned a score of 1 through 6 according to the following scheme: 1. eighth grade or less; 2. part high school; 3. high school graduate; 4. part college; 5. college graduate; 6. post graduate training.

Early Maternal Concern with Achievement and Development

For 44 of the 86 children (20 boys and 24 girls) there were lengthy reports, based on visits to the home, in which the interaction of mother and child was described. These reports were studied for the first three years of the child's life, and independent ratings were made by the authors for a variety of variables. All variables were rated on a seven-point scale. The variable relevant for this report was defined in terms of the degree to which the mother showed concern and anxiety over the child's cognitive and motor development and the degree to which she placed excessive expectations on the child's achievements. It reflected the degree to which the mother pushed the child's development beyond his abilities and her concern with his general achievement level. Major sources of evidence included concern over the age when the child walked, talked, rolled over, showing off of the child's skills and abilities and maternal dissatisfaction with the child's development. The interrater reliability was +.80 (product-moment correlation).

IQ Changes

Each child was given the Stanford-Binet, Forms L and M alternately, semiannually from ages 3 to 5 and annually from ages 6 through 11. All tests were administered by the same psychologist.[2] The mean IQ for the entire Fels population is about 120 (*SD*, 15). For each *S*, a smoothed plot of his IQ scores was obtained by averaging his three IQ scores around each age. For example, a child's smoothed or average IQ at age 6 was the result of averaging his IQ scores at

[2] The authors wish to acknowledge the assistance of Virginia L. Nelson who administered all of the IQ tests and who has been helpful during all phases of the research.

ages 5, 6, and 7; his smoothed IQ at age 10 was the average of his IQs at ages 9, 10, and 11. The procedure tends to remove the chance variation associated with any one IQ score and has been used in other studies (Kagan et al., 1958; Sontag, 1958). Each *S*'s smoothed IQ at age 6 was then subtracted from his smoothed IQ at age 10, and the resulting differences were used as the measure of IQ change.

RESULTS

Stability of Achievement Fantasy

There was a tendency for the proportion of *S*s reporting achievement themes to increase with age, and there were no marked sex differences at any one administration. Table 1 shows the percentage of each group reporting one or more achievement themes on each of the three protocols.

The average number of achievement themes per protocol was one, and Cards 1 and 17 BM, both suggestive of achievement plots, accounted for 76% of all achievement themes. Card 1 elicited more achievement themes than any card (43% of all themes) but, with age, there was an increase in the incidence of achievement themes to Card 17 and a decrease to Card 1. The typical achievement theme to Card 1 concerned a boy who wanted to master the violin and/or become a famous violinist, while the typical achievement story to Card 17 BM involved a person who was in a rope-climbing contest and wanted to do his best to win.

Since achievement fantasy was scored as either present or absent on any one protocol, the phi coefficient was used to express the reliability of the presence of the fantasy over the three administrations. Since there were no marked sex differences with respect to the reliability of the fantasy, the sexes were pooled in computing the final phi coefficients. Table 2 shows the value of phi and the associated one-tailed significance values for the intercorrelations among the three protocols. The occurrence of achievement fantasy on Protocol 1 was significantly associated with occurrence of the fantasy on Protocols 2 and 3.

TABLE 1
PERCENTAGE REPORTING ACHIEVEMENT THEMES ON THE THREE PROTOCOLS

Protocol	Boys	Girls	Boys and Girls
1	40.9	35.7	38.4
2	47.7	45.3	46.5
3	59.1	59.5	59.3

TABLE 2
RELIABILITY OF ACHIEVEMENT FANTASY FOR THREE PROTOCOLS
(Phi coefficients)

	Protocol 1	Protocol 2	Protocol 3
Protocol 1	—	+.32**	+.22*
Protocol 2		—	+.16

* $P < .05$.
** $P < .01$.

TABLE 3
FORMAL EDUCATION OF THE PARENTS
(Percentage of group with parents at each level of education)

Educational Level	Boys		Girls	
	Father	Mother	Father	Mother
1. Eighth grade or less	18.2	11.4	33.3	11.9
2. Some high school	9.1	13.6	9.5	11.9
3. High school graduate	20.5	31.8	16.7	26.2
4. Some college	6.8	22.7	21.4	26.2
5. College graduate	22.7	20.5	11.9	23.8
6. Post graduate	22.7	0.0	7.2	0.0

Educational Level of the Parents

In many families, one parent had more formal education than the other, and since the values of the same-sex parent were assumed to have special influence on the identification process in the child, analyses of this variable were kept separate for boys and girls. Table 3 summarizes the education of the 86 parent pairs under study.

The relation between educational level of the parents and occurrence of achievement fantasy on Protocol 1 was assessed by dichotomizing the variable of educational level into "Some college training" versus "No college training" and relating this dichotomy to presence or absence of achievement fantasy on the first protocol. Table 4 shows the resulting phi coefficients for boys and girls.

Children from homes in which both parents had college training were not much more apt to report achievement themes on Protocol 1 than children from homes in which neither parent had college training. The mean parental

TABLE 4

RELATION BETWEEN ACHIEVEMENT FANTASY ON PROTOCOL 1 AND EDUCATIONAL LEVEL OF PARENTS

(Phi coefficients)

	Education of Both Parents	Education of Father	Education of Mother
Boys	+.17	+.28*	+.11
Girls	+.09	−.11	+.01

* $P < .05$.

educational level of the group reporting achievement themes was 3.5, while the mean for the group without achievement fantasy was 3.2. Thus, achievement fantasy on the first protocol was not significantly related to the educational background of the family as a whole. However, for boys, there was a positive relation between education of the father and occurrence of achievement fantasy. The mean educational level of the fathers of boys with achievement themes was 4.7; the mean for the fathers of boys without achievement fantasy was 3.1 ($t = 3.20$; $p < .01$). In order to assess the stability of these results, phi coefficients were computed relating achievement themes on Protocols 2 and 3 with both the educational level of each parent and the family as a whole. For Protocol 2 there was no relationship between occurrence of achievement fantasy and the education of either or both parents. For Protocol 3, there was a positive relation, for girls only, between occurrence of achievement themes and maternal level of education (phi = +.26; $p < .05$).

In general, for this middle-class sample, level of parental education did not show any consistent or highly significant relationship to achievement fantasy in the child. However, the positive relation between achievement themes and educational level of the same-sex parent, which occurred for boys on Protocol 1 and for girls on Protocol 3, tentatively suggests that achievement concern on the part of the same-sex parent may have a special influence on the child's achievement strivings. This statement assumes that a highly educated parent would be more concerned with achievement in his child than a poorly educated parent.

Maternal Concern with Achievement

The mean ratings on maternal concern with achievement during the child's first three years were compared for *S*s with achievement fantasy present or absent on the first protocol. The mothers of girls who reported an achievement story were rated as more concerned with achievement than mothers of girls who did not report an achievement theme ($t = 2.65$; $p < .01$). The difference for the boys was negligible ($t = .79$; $p > .20$). The stability of this finding was assessed for Protocols 2 and 3, and similar results were obtained. For both Protocols 2 and 3, the mothers of the girls reporting an achievement theme were rated as more concerned with achievement than the mothers of the girls who did not report achievement fantasy ($p = .07, .05$ respectively). The differences for the boys were again negligible. Thus, the relation between early maternal concern with achievement and achievement fantasy in girls was relatively stable over a six-year period.

IQ Change and Achievement Fantasy

The dependent variable correlated with achievement fantasy was amount of increase in IQ over the ages 6–10. It should be noted that because of familiarity with the tests and the testing procedure, the average gain for a large sample of Fels children over this four-year period is 6 points and either no change or a loss in IQ is atypical for the average Fels *S*. Table 5 summarizes the data on base level IQ at age 6, mean IQ change and range of change for children high and low in achievement fantasy on the first protocol. There was no significant difference in base-level IQ at age 6 between the sexes, or between *S*s with achievement fantasy present or absent. However, *S*s who told an achievement theme on the first protocol showed a significantly larger gain in IQ (+10.7 points) than those who did not tell an achievement theme (+ 5.5 points) ($t = 2.63$; $p < .01$). In order to guarantee that this result was not a function of differences in IQ at age 6, an analysis of covariance was also applied to the data. The results revealed that IQ gain was significantly associated with the presence of achievement fantasy when variation due to base-level IQ at age 6 was partialed out ($F = 5.56$; $p < .01$).

TABLE 5
RELATION BETWEEN IQ GAIN AND ACHIEVEMENT FANTASY ON PROTOCOL 1

	Achievement Present			Achievement Absent		
	Boys	Girls	Both Sexes	Boys	Girls	Both Sexes
Mean IQ at age 6	120.2	116.2	118.4	110.2	115.4	112.9
Mean increase in IQ	11.6	9.7	10.7	7.4	3.7	5.5
Range of IQ changes	−12 to +40	+1 to +24		−11 to +26	−11 to +16	

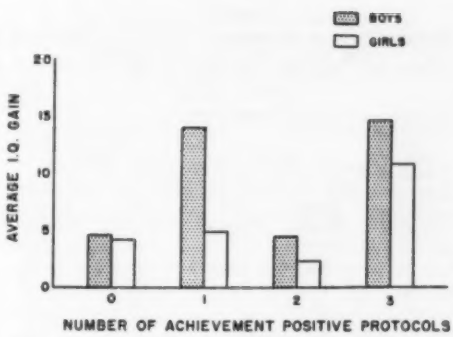


FIG. 1. RELATION BETWEEN IQ GAIN AND NUMBER OF ACHIEVEMENT POSITIVE PROTOCOLS

Figure 1 illustrates the relation between the amount of IQ gain and the number of protocols in which one or more achievement themes occurred (achievement positive protocols). Evidently the relationship is not linear. Although *S*s with three achievement positive protocols showed a greater gain in IQ than *S*s having no achievement positive protocols ($t = 2.74$; $p < .01$), the *S*s with two positive protocols showed less IQ gain than those with one positive protocol ($t = 2.05$; $p < .05$). The drop in IQ gain for *S*s with two achievement protocols was more marked for boys than for girls for the difference in IQ gain between *S*s with one versus two positive protocols was 9.5 points for the boys and only 2.6 points for girls.

If one regards number of achievement positive protocols as an index of the strength of achievement strivings, then this result is in agreement with other research which has shown a nonlinear relation between achievement themes and a class of overt achievement behavior (Clark, Teevan, & Ricciuti, 1956; McClelland et al., 1953). McClelland has hypothesized that *S*s who show a moderate amount of achievement fantasy are characterized by both a strong desire for achievement goals and a strong fear of failure and assumes that the fear of failure elicits responses that oppose and block successful achievement behavior. In an attempt to ascertain whether fear of failure fantasy was related to amount of IQ change the mean IQ gain of the 25 *S*s who reported one or more fear of failure themes on any one of the protocols was compared with the mean gain of the 43 *S*s who only reported positive, success achievement themes. The mean gains were 9.7 and 9.4 points respectively indicating no marked difference between these two groups. Thus, for this sample, fear of failure fantasy was not related to inferior performance on the IQ increase measure.

Another possibility is that the testing situation was not a strong incentive for all the *S*s, and only those children who sought gratification of achievement goals in an intellectual area would strive to perform well on the intelligence test. This hypothesis was supported by information on the elementary school grades of 35 boys and 37 girls from the original sample of 86. Each child's average grade over the four-year period covering elementary grades 2–5 was computed. Then the boys and girls were ranked separately with relation to their mean grade and each distribution dichotomized at the median into high and low grade groups. Table 6 shows the distribution of achievement positive protocols for boys and girls above or below the median in school grades.

For girls there was no difference between the high and low grade groups with respect to the distribution or number of achievement-positive protocols. However, the boys with two achievement-positive protocols were more likely to have low grades than those with one such protocol ($p < .01$; exact test). This finding may help to clarify the meaning of the minimal IQ gain shown by the boys with two achievement-positive protocols. That these same boys performed poorly in school suggest that an intellectual task was not a strong incentive for them. It is tentatively suggested that their achievement strivings were directed toward goals related to nonintellectual activities like athletics or manual skills.

TABLE 6
DISTRIBUTION OF ACHIEVEMENT POSITIVE PROTOCOLS FOR BOYS AND GIRLS ABOVE AND BELOW THE MEDIAN IN SCHOOL GRADES
(Percentages)

No. of Achievement Positive Protocols	Boys		Girls	
	High Grades	Low Grades	High Grades	Low Grades
0	23.5	22.2	15.0	23.5
1	47.0	16.7	35.0	35.3
2	6.0	50.0	20.0	17.7
3	23.5	11.1	30.0	23.5
Total	100.0	100.0	100.0	100.0

These results seem to indicate that gain in IQ during childhood is one possible measure of achievement-strivings with respect to intellectual activity. The product-moment correlations between average school grade, for the grades 2–5, and IQ gain was +.48 for girls and +.25 for boys ($p < .01$, $< .10$ respectively). In addition, amount of IQ gain does not seem dependent on initial intellectual capacity, for the product-moment correlation between Binet IQ at age 3 and amount of IQ change was nonsignificant ($r = +.15$), while the correlation between IQ at age 3 and at age 6 was +.73 ($p < .01$).

Maternal Concern with Achievement and IQ Change

Since early maternal concern with achievement was positively correlated with achievement fantasy for girls, and achievement fantasy correlated with IQ gain, it was suspected that early maternal concern with achievement would be positively correlated with IQ gain in girls but not necessarily with boys. These assumptions were verified, for the product-moment correlation for girls was +.37 ($p < .05$), while for boys it was −.10 ($p > .20$).

DISCUSSION

The results indicate that achievement fantasy shows more than chance stability during childhood and early adolescence. The data, therefore, add validity to predictive and explanatory statements about achievement behavior derived from fantasy measures.

With respect to antecedent factors, it seems that for this select middle-class sample achievement fantasy was not significantly related to the general level of education in the family but partly dependent on the achievement concern of the same-sex parent, for maternal concern with achievement correlated only with achievement themes in girls. It is suggested that concern with achievement behavior and the sex of the parent showing the concern may interact in influencing the acquisition of achievement-strivings in the child.

At a more speculative level, it is possible that an opposite-sex parent's concern with achievement may lead to some inhibition of achievement-striving in the child if he perceives achievement behavior as an attribute of the opposite sex. Such speculations agree with psychoanalytic hypotheses to the effect that learning problems often arise when the child associates intellectual achievement with the goals and attributes of the opposite-sex parent (Blanchard, 1946; Pearson, 1952), and the consequent anxiety over sex role identity inhibits learning and intellectual effort.

There are several cautions with respect to generalizing from these data. First, the specific fantasy stimuli used tended to elicit themes emphasizing success and the acquisition of achievement goals, rather than failure or anxiety over not attaining them. Clark, Teevan, and Ricciuti (1956) have reported that the relationships between achievement fantasy and a criterion of overt achievement were quite different when the fantasy stimuli suggested success stories from when they suggested failure. Second, the stimulus that accounted for most of the achievement themes was Card 1 (boy with violin), which typically evoked the theme of intellectual mastery of the instrument. It is possible that if there were no fantasy stimulus suggesting achievement on an intellectual task the relationship with IQ gain and level of parental education might not have appeared. It is suggested that the greater the comparability in content between the fantasy and overt behavior, the greater the predictive power of fantasy.

The occurrence of any achievement-related behavior, be it IQ gain or a TAT theme, it is suggested, is a result of the interaction of (*a*) the intensity of the motivation for a specific achievement goal, (*b*) the strength of response

tendencies directed toward attaining it, and (*c*) the strength of response tendencies that oppose striving toward attaining the goal. Motive is defined, after McClelland, as the anticipation of a specific goal. However, this anticipatory state does not imply any specific cognitive or instrumental response associated with it, and the specific behaviors that follow arousal of a motive depend, in part, on the strength of goal-directed responses and the strength of anxiety-based avoidant responses associated with the motive. Two potential sources of anxiety that may be associated with achievement goals are fear of failure and association of intellectual achievement with opposite-sex attributes. The strength of these anxiety-based avoidant responses may interact with the approach-response tendencies in influencing the occurrence of both direct achievement behavior and achievement fantasy. Achievement fantasy is, therefore, viewed as one index of the strength of the response tendency to strive for a specific achievement goal.

Recent research (Kagan, 1956; Mussen & Scodel, 1955) has seriously questioned the validity of the prevalent assumption that a projective stimulus minimizes the effects of anxiety on the expression of a motive and thus provides an index of "pure" motive strength. Since the stimuli used to obtain achievement themes usually suggest achievement content it seems plausible that if an *S* has strong anxiety over achievement goals, achievement themes may not appear in the protocol despite strong achievement motivation. At present there are no well-validated techniques that allow an investigator to use various aspects of the person's contemporaneous behavior to assess motivation as well as approach and avoidant response tendencies. One solution to this problem may lie in the use of antecedent, historical data as indices of motivation.

SUMMARY

The following data on 44 boys and 42 girls in the Fels research population were studied to assess the reliability and validity correlates of achievement fantasy.

a. Three TAT protocols administered at ages 8-9, 11-6, and 14-6.

b. Educational level of the parents.

c. Ratings of maternal concern with achievement for the first three years of the child's life.

d. Changes in IQ score during the years 6 through 10.

The results showed (*a*) an increase in occurrence of achievement themes with age and better than chance stability of occurrence over time, (*b*) a positive relation between early maternal concern with achievement and both achievement fantasy and IQ gain in girls, (*c*) a nonlinear but generally positive relation between occurrence of achievement fantasy and IQ gain for boys and girls, and (*d*) a suggestive relation between achievement fantasy and educational level of the same-sex parent.

It was suggested that achievement fantasy is an index of the strength of the subject's behavioral tendency to seek achievement goals and that achievement concern on the part of the same-sex parent has a special influence on the child's achievement strivings. Specific methodological and theoretical problems with respect to measurement and interpretation of achievement fantasy were discussed.

REFERENCES

ATKINSON, J. W. Explorations using imaginative thought to assess the strength of human motives. In M. R. Jones, *Current theory and research in motivation: A symposium:* Lincoln: Univer. Nebraska Press, 1954. Pp. 56–112.

ATKINSON, J. W. Motivational determinants of risk-taking behavior. *Psychol. Rev.*, 1957, **64**, 359–372.

BLANCHARD, PHYLLIS. Psychoanalytic contributions to the problems of reading disabilities. In *Psychoanalytic study of the child.* II. New York: Int. Univer. Press, 1946. Pp. 163–188.

CLARK, R. A., TEEVAN, R., & RICCIUTI, H. N. Hope-of-success and fear-of-failure as aspects of need for achievement. *J. abnorm. soc. Psychol.*, 1956, **53**, 182–186.

FARBER, I. E. Comments on Professor Atkinsons' Paper. In M. R. Jones, *Current theory and research in motivation: A symposium.* Lincoln: Univer. Nebraska Press, 1954. Pp. 112–115.

FRENCH, ELIZABETH G. Some characteristics of achievement motivation. *J. exp. Psychol.*, 1955, **50**, 232–236.

FRENCH, ELIZABETH G. The interaction of achievement motivation and ability in problem-solving success. *J. abnorm. soc. Psychol.*, 1958, **57**, 306–309.

FRENCH, ELIZABETH G., & THOMAS, F. H. The relation of achievement motivation to problem-solving

effectiveness. *J. abnorm. soc. Psychol.*, 1958, **56**, 45–48.

KAGAN, J. Measurement of overt aggression from fantasy. *J. abnorm. soc. Psychol.*, 1956, **52**, 390–393.

KAGAN, J., SONTAG, L. W., NELSON, VIRGINIA L., & BAKER, C. T. Personality and IQ change. *J. abnorm. soc. Psychol.*, 1958, **56**, 261–266.

LAZARUS, R. S., & BAKER, R. W. Motivation and personality in psychological stress. *Psychol. Newsltr., N.Y.U.*, 1957, **8**, 159–193.

LOWELL, E. L. A methodological study of projectively measured achievement motivation. Unpublished master's thesis. Wesleyan Univer., 1950.

MCCLELLAND, D. C., & LIBERMAN, A. M. The effect of need for achievement on recognition of need-related words. *J. Pers.*, 1949, **18**, 236–251.

MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive.* New York: Appleton-Century-Crofts, 1953.

MUSSEN, P. H., & SCODEL, A. The effects of sexual stimulation under varying conditions on TAT sexual responsiveness. *J. consult. Psychol.*, 1955, **19**, 90.

PEARSON, G. H. J. A survey of learning difficulties in children. In *Psychoanalytic study of the child.* VII. New York: Int. Univer. Press, 1952. Pp. 322–356.

SONTAG, L. W., BAKER, C. T., & NELSON, VIRGINIA L. Mental growth and personality development. *Monogr. Soc. Res. Child Developm.*, 1958, **23**, No. 68.

VOGEL, W., LAZARUS, R. S., & BAKER, R. W. A comparison of fantasy and behavioral indices as predictors of extrinsically motivated behavior under conditions including a failure stress. *Amer. Psychol.*, 1957, **12**, 400.

WINTERBOTTOM, M. R. The relation of childhood training in independence to achievement motivation. Unpublished doctoral dissertation, Univer. of Michigan, 1952.

Received February 7, 1958.

NEED AFFILIATION AND FREQUENCY OF FOUR TYPES OF COMMUNICATION

JOHN B. LANSING AND ROGER W. HEYNS[1]

University of Michigan

THIS paper reports an effort to predict social behavior from a personality variable. Specifically, it is another study in a series aimed at discovering the behavioral correlates of need for affiliation. Shipley and Veroff (1952) developed a projective measure of need for affiliation, defining the motive primarily in terms of separation anxiety and concern over rejection. Atkinson, Heyns, and Veroff (1954) in a similarly designed study constructed a measure based upon a more positive definition: *concern with the establishment, maintenance, or restoration of a code of personal relationship, best symbolized by the word friendship.* French and Chadwick (1956) have recently reported results that are consistent with this more positive conception of affiliation motivation. It is this definition which is used in the present study.

Data concerning the empirical correlates of the need for affiliation are still sparse, since the methods of measurement are recent. French (1956), using a somewhat different measure but the same conception of need for affiliation, has reported results which indicate that the subject's choices of a work partner are correlated with need for affiliation and need for achievement. Subjects with high need for affiliation and low achievement motivation pick a less competent friend for a work partner, while subjects with high need for achievement and low affiliation motivation choose a competent nonfriend. In the validation study, Atkinson et al. (1954) report that the persons with the highest need for affiliation were more likely to be cited by their friends as seeking approval. Both the Shipley-Veroff (1952) and the Atkinson et al. (1954) studies report a negative correlation between need for affiliation and popularity in the group. The present study is an attempt to extend our knowledge of the relationship between the need for affiliation and other social acts.

The data reported here were collected in an investigation of motives, attitudes, and long distance telephone calls. In the study, a number of independent variables were measured which might be related to use of the long distance telephone for social calls; several alternative channels of communication were also investigated. The present report is concerned with the relationship between n Affiliation and frequency of use of these four forms of communication: (*a*) use of long distance telephone for social calls, (*b*) use of the local telephone, (*c*) writing letters, and (*d*) visiting close friends and relatives living at a distance.

The initial hypothesis held that people who score high on n Affiliation also tend to be frequent users of each of these methods of communication, other things being equal. Among the relevant other factors to be controlled (discussed below) is, for example, income: a person's n Affiliation should be expected to predict the frequency of his social long distance calling only when his income is taken into account.

This research is unusual in that it attempts to quantify the relation between a measure of personality and the behavior of a segment of the general population. Published studies known to us for the most part are concerned with more restricted populations, commonly populations of college students. Also, the present study differs from laboratory investigations in that the behavior measured is behavior in "real life," uninfluenced by the investigation and measured after the event. The methods used here have their difficulties and are not necessarily superior. They are, however, of interest as a complement to other techniques. One objective of the study of personality, in our view, is to measure dimensions of personality that can be related quantitatively to the behavior of the population at large. We all tend to apply research results to

[1] The authors wish to acknowledge their indebtedness to their colleagues at the Survey Research Center, particularly to Charles Cannell, Joseph Adelson, and John Colombotos who were closely associated with the larger research project that is reported here in part.

the general population in our own thinking. It seems to us that there is a need for research that attempts systematically to carry through this process of generalization.

METHOD

The sample used in this research was designed especially for the study of frequency of use of residential telephones for long distance calls. The universe sampled consisted of residential subscribers of the Michigan Bell Telephone Company and their wives, in the southern part of the lower peninsula of the state of Michigan. A sample of 1571 subscribers was drawn, and information was tabulated from company records about the frequency of use of long distance by each. For purposes of this study, a "long distance" call was defined as a call to a point over 50 miles from the point of origin. At this point three groups were defined, "low" users of long distance, "intermediate" users, and "high" users. A "high" user is defined to be a subscriber who made six or more long distance calls in the last period of three months for which data were available at the time when the sample was selected (in the early fall of 1955). A "low" user was a subscriber who made no long distance calls in the same period. Personal interviews were taken with "low" users and "high" users but not with "intermediate" users. Although the number of low users in the population is much larger than the number of high users, half the interviews were taken with low users and half with high users, to permit efficient comparison of the groups.

Once a subscriber was selected for interview, if he was married his wife also was designated to be interviewed. If the subscriber had no spouse, only the subscriber fell in the sample. Interviews were taken with 400 individuals representing 230 families.

This sample design, using as it does extremes in the behavior of immediate interest, is efficient for the investigation of factors related to the frequency of long distance calling. It is of uncertain efficiency for the investigation of factors relating to the frequency of other forms of communication. Estimates of relationships between independent variables and these other types of communication will be in error if the independent variables enter into different relationships among people who are intermediate in their use of long distance from those that are characteristic of high and low users. We have no reason to believe that the effect of the independent variables is complicated in this manner.

The sampling procedure does not permit estimation of the frequency of different types of communication among the population as a whole. For example, the data should not be used to estimate the mean number of letters written per year per adult in Michigan.

The interviewing techniques were similar for the most part to those used in other cross-section studies by the Survey Research Center (Campbell & Katona, 1953). The interviews lasted about an hour. A common pattern was for the interviewer to interview the wife during the day and return in the early evening to interview the husband.

The Affiliation motive was measured by means of responses to TAT pictures. Each respondent was shown a series of four pictures and asked to tell a story about each.[2] The general procedure followed here has been described in detail elsewhere (Atkinson et al., 1954). Some of the pictures that had been satisfactory with college students were not successful, largely because the persons in the pictures were too youthful.[3] The interviewers received special instruction in the use of the pictures, and had no serious difficulty in using them in the interview. Training was directed particularly to overcome a tendency observed in the pretest for protocols to be too brief to be scored properly. The scoring of the protocols was done by the second author, according to the published manual (Heyns, Atkinson, & Veroff, 1958). Before scoring the present material he checked his reliability by scoring 50 stories that had been set aside for purposes of training and reliability assessment. The total scores he assigned to the 50 stories correlated .94 with those which had previously been ascribed to them.

The techniques used in the analysis of the data include the familiar frequency tabulations and multiple regression. The estimates of sampling error for the regression coefficients shown were calculated by a technique that assumes that the sample used was a simple random sample. In fact, the sample is clustered. That is, the units in the sampling were not selected independently.

The problems of estimating sampling errors of clustered samples have been discussed elsewhere (Kish, 1957). The essential difficulty is that the usual formulas lead to underestimates of the true sampling errors when applied to clustered samples such as the present one. The drawing of the present sample proceeded in stages. In the interests of economy, interviews were taken only in certain telephone exchanges in southern Michigan instead of being spread all over that part of the state. Our sample was large enough so that we took interviews in all metropolitan areas, and these covered about two-thirds of our population. As a result, though we do not have quite as much information as we would if the interviews had been spread more widely, the usual formulas for estimating the sampling

---

[2] The sequence of questions used was as follows:

I'm going to show you some pictures and I'd like you to make up a story about them.
(*a*) What's going on here?
(*b*) What is the person (what are they) thinking?
(1) What do they want?
(*c*) What's going to happen next?

[3] The pictures finally chosen were as follows: For women (*a*) a young girl standing in a transportation terminal, (*b*) three young women talking over tea cups, (*c*) two men and two women around a desk in an office, and (*d*) huddled figure of young person with head on a cot. For men (*a*) two men, one sitting on a fence, rural background, (*b*) a male passenger on an airplane, (*c*) four men sitting and lounging informally in a furnished room, and (*d*) two men standing, facing each other, talking.

TABLE 1

RELATIONSHIP BETWEEN NEED AFFILIATION SCORE AND NUMBER OF LOCAL CALLS MADE OR RECEIVED "LAST" WEEK[a]

Number of Calls	All Respondents	Need Affiliation Score		
		Low (00-03)	Middle (04-06)	High (07-19)
Up to 2	18	21	21	10
3 to 14	49	52	51	45
15 and over	29	22	26	41
Not ascertained	4	5	2	4
Total	100	100	100	100
Number of respondents	400	155	103	125

[a] The question was: "About how many calls in the last week have you yourself made or received on your home phone?" The question is phrased to include all types of calls, but answers may be assumed to refer primarily to local calls.

TABLE 2

RELATIONSHIP BETWEEN NEED AFFILIATION SCORE AND NUMBER OF TIMES A YEAR RESPONDENT GETS TOGETHER WITH OUT-OF-TOWN FRIENDS OR RELATIVES[a]

Number of Visits	All Respondents	Need Affiliation Score		
		Low (00-03)	Middle (04-06)	High (07-19)
Up to 1	45	51	49	36
1 to 10	36	31	36	39
11 or more	10	9	6	14
Other answers or not ascertained	9	9	9	11
Total	100	100	100	100
Number of respondents[b]	308	118	80	101

[a] The question was: "Thinking of your half dozen closest friends or relatives, do they all live here in (name of city), or do they live in other places, or what? About how often do you get together with these people?"

[b] Number of respondents includes only those respondents who have relatives or friends living 50 miles or more away.

error of proportions can be used with minor adjustments.[4]

A more serious complication arises because both husband and wife were interviewed in the family. If husband and wife behave as a unit with respect to the behavior under study, we do not have as many independent observations as we have interviews. In fact, we have only as many independent observations as we have married couples in the sample. If husband and wife behave independently, however, the number of independent observations is the same as the number of interviews. Thus, where there is no correlation between data from the interview with the husband and data from the interview with the wife, the number of interviews is equivalent to the number of observations and should be used in entering the table above. When perfect positive correlation exists, the number of observations is equal to the number of married couples.

Regression analysis requires that the variables used be scaled. As will be apparent from a study of the variables used, some of the scales imposed on the data were defined by the use of judgment. To the extent that the judgment used was poor, the effect is to reduce the chances of obtaining significant results. Operationally, our hypotheses are to the effect that the independent variables as scaled will have an effect on the dependent variables as scaled.

[4] Differences equal to or greater than those shown below may be taken as significant at the 95% level:

Approximate Value of the Percentage Found	Minimum Difference Between Two Groups Needed for Significance	
	For two groups each of 200	For two groups each of 100
35 to 65%	11	15
20 or 80%	9	12
10 or 90%	7	—

RESULTS

Frequency Analysis

Tables 1 to 4 present the basic data concerning the relationship of n Affiliation to frequency of each of four kinds of communication in turn.

The relation between n Affiliation score and number of local calls made or received by the person "last" week (Table 1) is statistically significant and in the predicted direction. There is no necessary correlation between the number of calls on the home phone by husband and wife, but we may allow for clustering by conservatively estimating that we have 100 low and 100 high scores on n Affiliation instead of the 155 and 125 observed. We require for statistical significance a difference of 14 percentage points between the proportion of low scorers making 15 or more calls and the proportion of high scorers making 15 or more calls. The obtained difference is 19 percentage points.

In the next three tables, which concern communication with friends and relatives living 50 miles or more away, we omit respondents who report that none of their half dozen closest friends and relatives live at that distance.

Table 2 shows a relationship between n

TABLE 3

RELATIONSHIP BETWEEN NEED AFFILIATION SCORE AND NUMBER OF LETTERS WRITTEN PER YEAR TO RELATIVES OR FRIENDS[a]

(Percentage distribution of respondents)

Number of Letters	All Respondents	Need Affiliation Score		
		Low (00-03)	Middle (04-06)	High (07-19)
Up to 20	56	56	55	52
21 to 75	18	20	16	18
76 or more	21	15	25	26
Other answers or not ascertained	5	9	4	4
Total	100	100	100	100
Number of respondents[b]	308	118	80	101

[a] The question was: "Thinking of your half dozen closest friends and relatives, do they all live here in (name of city), or do they live in other places, or what? About how many letters would you say you write to relatives or friends in a year?"

[b] Number of respondents includes only those respondents who have relatives or friends living 50 miles away or more.

TABLE 4

RELATIONSHIP BETWEEN NEED AFFILIATION SCORE AND FREQUENCY OF USE OF LONG DISTANCE

(Percentage distribution of respondents)

Frequency of Calling	All Respondents	Need Affiliation Score		
		Low (00-03)	Middle (04-06)	High (07-19)
Low Users	43	37	50	46
High Users	57	63	50	54
Total	100	100	100	100
Number of respondents[a]	308	118	80	101

[a] Number of respondents includes only those respondents who have relatives or friends living 50 miles or more away.

Affiliation score and number of visits that is in the predicted direction but too small for significance if we allow for clustering.

Table 3 shows a relationship between n Affiliation score and number of letters written per year that is also below the margin of statistical significance but in the predicted direction. Table 4 shows no relation between n Affiliation score and frequency of use of long distance which is probably the most complexly determined of the four types of communicative acts. The interpretation of this result is discussed in some detail later.

Thus, our two-way tabulations show a complicated picture. There is a clearly significant relation between n Affiliation and the use of the local phone. The relationship of the need to frequency of letters and visits is in the predicted direction but is not significant. Finally, there is no relationship between the need measure and use of the long distance phone. We turn to more powerful analytical methods.

Regression Analysis

Tables 5 and 6 define the variables used in regression analysis of visits and frequency of letter writing, respectively. The independent variables in the two equations are identical. Family income (X_1) was introduced on the hypothesis that high income people are more likely to make visits at a distance and to write letters. X_2 is a measure of the closeness (degree of kinship) of the relatives living away. The hypothesis is that those with only distant relatives away are less frequent visitors and less prolific letter writers than those whose close relatives live away. Distance (X_3) forms a barrier to visits, and, we predicted, may indirectly weaken all ties, reducing even letter writing. Education (X_4) as a measure of socioeconomic status may have an effect on visiting, on the hypothesis that people with high education travel more, even for a given income level. Education, as a measure of familiarity with the written word, should be related to frequency of letter writing. Occupation (X_5) was introduced as an additional measure of socioeconomic status, with the direction of prediction similar to that for education. Need Affiliation scores ranging from 0 to 19 enter as X_6.

The results for visits are as follows:

$$[1] \quad Y_1 = 3.44 + \underset{(.04)}{.05X_1} - \underset{(.06)}{.02X_2} - \underset{(.05)}{.29X_3} + \underset{(.06)}{.01X_4} + \underset{(.06)}{.11X_5} + \underset{(.025)}{.049X_6}$$

$$R = .363$$

$$R^2 = .13$$

Under each coefficient, in parenthesis is its standard error (one sigma). Our interest is primarily in n Affiliation (X_6), which shows a coefficient equal to two times its standard error and in the predicted direction. As noted above, the sample design is such that the

TABLE 5

DEFINITION OF VARIABLES FOR REGRESSION ANALYSIS OF FREQUENCY OF VISITS, EQUATION [1]

Variable	Symbol	Range of Values
Frequency of visits	Y_1	1. Less than once a year 2. Once a year; "very irregularly" 3. 1–5 times a year; frequency of visits not ascertained 4. 6–10 times a year 5. 11–15 times a year; "fairly frequently," "regularly" 6. 16 or more times a year
Family income	X_1	1. Under $1000 2. $1000–1999 3. $2000–2999 4. $3000–3999 5. $4000–4999; income not ascertained 6. $5000–5999 7. $6000–7499 8. $7500–9999 9. $10,000 or more
Closeness of relatives living away	X_2	0. Has no relatives living away (Interviews coded 0 on this scale were not included in the regression) 1. Has only distant relatives away 2. Has both close and distant relatives away 3. Has only close relatives away
Distance friends and relatives live away	X_3	1. 50–299 miles (midpoint 175 miles) 2. 50–299 and 300–600 (midpoint 312 miles) 3. 300–600 only; 50–299 and over 600; not ascertained (midpoint 470 miles) 4. 300–600 and over 600 (midpoint 625 miles) 5. Over 600 miles (midpoint 800 miles)
Education of head of family	X_4	1. Attended grammar school (1–8 yrs.) 2. Attended high school (9–11 yrs.); 1–8 yrs. plus other education 3. Completed high school; education not ascertained 4. Attended college, no degree 5. College degree
Occupation of head of family	X_5	1. Laborer, service worker 2. Craftsman, foreman, operative, member of armed forces 3. Housewife, retired; occupation not ascertained 4. Clerical, sales 5. Professional, managerial, self-employed
Need affiliation	X_6	00. – 19. (actual score)

TABLE 6

DEFINITION OF DEPENDENT VARIABLE FOR REGRESSION ANALYSIS OF FREQUENCY OF LETTER WRITING,[a] EQUATION [2]

Variable	Symbol	Range of Values
Frequency of letter writing	Y_2	0. None at all 1. 1–5 letters; "just a few" 2. 6–20 letters 3. 21–40 letters, number not ascertained 4. 41–50 letters 5. 51–75 letters 6. 76 letters or more

[a] The independent variables in Equation [2] are identical with those in Equation [1]. See Table 5.

standard errors shown are underestimates of the true errors. If we were dealing with a simple random sample, the probability of such an event occurring by chance would be about .03. We may look at the matter in another way by saying that a departure of 1.65 sigmas from zero in a predicted direction will occur by chance 5% of the time. Our departure from zero is .049. If the true sigma is about .03 or smaller, our result is significant at the 5% level. We suspect that the effect of clustering is powerful enough to raise the standard error from .025 to more than .03. We can say only that the data seem to point in the direction indicated by the hypothesis.

In interpreting the results for the other variables we may consider jointly X_1 (income), X_4 (education), and X_5 (occupation). These variables are intercorrelated, with simple correlation coefficients ranging from .37 (X_1 with X_4) to .58 (X_4 with X_5). In view of these two-way coefficients, it is not surprising that no one of the variables has a significant incremental effect. The distance people live away (X_3) is shown to have an effect on the frequency of visits. Degree of kinship of relatives (X_4) has no effect.

The results for letters are as follows:

$$[2]\quad Y_2 = 2.19 + \underset{(.07)}{.01X} - \underset{(.11)}{.19X_2} - \underset{(.09)}{.10X_3} + \underset{(.11)}{.10X_4} + \underset{(.11)}{10X_5} + \underset{(.36)}{.84X_6}$$

$$R = .126$$
$$R^2 = .02$$

Again, we are interested primarily in n Affiliation (X_6). Again, the results would be significant without question if we were dealing with a simple random sample, since the coefficient of .84 is more than twice its standard error (.36) and is, furthermore, in the direction predicted by the theory. We have no reason to believe that husbands and wives write letters jointly as commonly as they visit distant relatives together. In fact, we would anticipate that as often as not if one of the two is an enthusiastic letter writer the other may write fewer letters. Hence, it seems to us unlikely that the true standard error is enough larger than .36 to change our conclusions. (An error of .51 or above would be needed.) Thus, as best we can judge, the data support the hypothesis. We note, however, that the multiple correlation coefficient is only .126, and its square, only .02.

Of the other independent variables considered, none show any effect. Only X_3 (closeness of relative living away) approaches significance at all, and here the sign of the coefficient is the reverse of that predicted.

Finally, a multiple regression was computed, not presented here, using as dependent variable whether the individual (or spouse) was a low user or a high user of long distance and using score on n Affiliation as one of the independent variables. The results as far as n Affiliation is concerned were as unpromising as those in Table 4.

We did not compute a multiple regression for local telephone calls. As a by-product of the calculations described above we computed two-way correlation coefficients between each pair of independent variables, and found none of our demographic variables to be correlated with n Affiliation. Thus, there was no possibility that the correlation observed in Table 1 could be attributed to an association between n Affiliation and one of the demographic variables even if one of the latter proved to be related to the number of local calls.

DISCUSSION AND FURTHER ANALYSIS

The analysis up to this stage left the investigators with an unresolved problem. Need Affiliation as measured predicts local calling and letter writing and may be related even to visiting people at a distance. Why, then, does it not predict social long distance calling?

One possible answer is that n Affiliation is not relevant at all to long distance telephoning. We have tested this hypothesis by examining the relation between n Affiliation and other scores that bear on long distance calling. One of these was a measure of respondents' feelings while making long distance calls (Table 7). A second score was obtained from answers to a direct question as to whether people enjoy making long distance calls (Table 8).

The data show little relation between n Affiliation and the score on feelings while making a call. (This score, it should be added, *is* related to frequency of use of long distance.) The data do, however, show a relation between his score on n Affiliation and whether a respondent says he enjoys making long distance calls. From these data we conclude that n Affiliation as measured is not totally irrelevant to use of the long distance phone. At a minimum it does have something to do with people's reports of whether they enjoy making long distance calls.

There are several possible explanations of why n Affiliation might be relevant and yet show no direct relationship to frequency of long distance in our data. The first hypothesis is the

TABLE 7

RELATIONSHIP BETWEEN NEED AFFILIATION SCORE AND SCORE ON RESPONDENT'S FEELINGS WHILE PLACING A LONG DISTANCE CALL[a]

Feelings	All Respondents	Need Affiliation Score		
		Low (00–03)	Middle (04–06)	High (07–19)
Pleasant feelings	59	58	54	69
About half and half—neutral	23	25	26	15
Unpleasant feelings	11	13	13	9
Not ascertained	7	4	7	7
Total	100	100	100	100
Number of respondents[b]	347	132	87	115

[a] The question was: "Here is a list of ways people tell us they feel when they make a long distance call. Do any of these words describe how you sometimes feel when making a long distance call? Relaxed? Nervous? Tense? Anxious? Happy?" Scoring involved assigning +1 to each answer indicating presence of a positive feeling or absence of a negative feeling.

[b] Does not include those respondents who have never made a long distance call.

TABLE 8
RELATIONSHIP BETWEEN NEED AFFILIATION SCORE AND FEELINGS ABOUT MAKING LONG DISTANCE CALLS[a]
(Percentage distribution of respondents)

Feelings	All Respondents	Need Affiliation Score		
		Low (00-03)	Middle (04-06)	High (07-19)
Enjoy it; enjoy it moderately	60	55	61	70
Sometime enjoy, sometime dislike; neither enjoy nor dislike	27	33	25	22
Do not enjoy	6	7	6	1
Not ascertained	7	5	8	7
Total	100	100	100	100
Number of respondents[b]	347	132	87	115

[a] The question was: "How do you feel about making a long distance call—do you enjoy it or dislike it or what?"
[b] Does not include those respondents who have never made a long distance call.

simplest: that the effect of n Affiliation is too small to detect, using our data. In support of this contention one can call attention to the small size of the sample and the difficult problems of measurement. Weaknesses in the n Affiliation measure would be even more serious if the affiliative motive has only a slight impact. One may note further that the available data from the telephone company refer to the total number of calls placed per *telephone*, while the n Affiliation scores are for *individuals*. A variation on this argument is to take the position that only those who score very high on n Affiliation are different from the rest of the population. According to this view, most of the observed variation in affiliation score is irrelevant. And one may recall that the observed effects of n Affiliation on visiting and letter writing are at best small. All of these points seem to us to add up to a plausible case that there is an effect, but it is too small for us to detect.

A second hypothesis is more complex. It is that the effect of n Affiliation depends in some complex way upon the effect of other variables. The need for affiliation, according to this view, has its predicted effect only in certain circumstances; indeed, its effect may even be the reverse from one set of circumstances to another. Or, it may be associated with other variables whose effect varies from one set of circumstances to another.

Another possible area of speculation is that of operation of the affiliative motive itself. There is a clear relationship of the motive measure to communication acts such as local calls, letter writing, and a very suggestive one to visits. These acts may be successful enough in the lives of our respondents to satisfy the affiliative need, leaving only distance calling to be influenced by other factors. One would expect a relationship between the motive measure and long distance calls if these other acts were unsuccessful.

A fourth possibility concerns the measure itself. The definition of affiliative imagery and the scoring instructions are oriented toward friendship as the prototype of the affiliative relationship. Statements of family ties are not automatically scored as affiliative; there must be evidence of concern about the relationship. This is a stringent requirement; it means that the measure is less slanted toward ties between relatives. Our data show, however, that social long distance calls are primarily to relatives. It may be that a measure more sensitive to the desire for close ties to relatives would show a greater relationship. The same reasoning would argue that the present measure might predict amount of calling to friends. Unfortunately, it was not possible to explore this possibility.

This discussion has emphasized the failure of the Affiliation measure to predict social long distance calling because of the intriguing theoretical and methodological issue raised by the missing relationship. It is important to note, however, that this kind of speculation is a reasonable activity only because of our most important finding, that the n Affiliation measure is significantly related to some other communication acts, particularly local calling and letter writing and perhaps to making visits to friends.

SUMMARY

This article reports the results of an attempt to relate need for affiliation to the frequency of four types of communication: (*a*) use of long distance telephone for social calls, (*b*) use of the local telephone, (*c*) writing letters, and (*d*) visiting close friends and relatives living at a

distance. The data were collected in 400 personal interviews with a specially designed sample of telephone subscribers living in southern Michigan. The research is unusual in that it concerns the relation between a dimension of personality, measured by TAT-type pictures, and the behavior of a segment of the general population.

The major findings are that n Affiliation is correlated in the predicted direction with use of the local telephone. The same relationship is found to letter writing, though the correlation is low. With regard to visits, the correlation is in the predicted direction but not significant. Need affiliation as measured does not predict social long distance calling.

REFERENCES

ATKINSON, J. W., HEYNS, R. W., & VEROFF, J. The effect of experimental arousal of the affiliation motive on thematic apperception. *J. abnorm. soc. Psychol.*, 1954, **49**, 405–410.

CAMPBELL, A. A., & KATONA, G. The sample survey. In L. Festinger & D. Katz (Eds.), *Research methods in the behavioral sciences*. New York: Dryden, 1953.

FRENCH, ELIZABETH G. Motivation as a variable in work partner selection. *J. abnorm. soc. Psychol.*, 1956, **53**, 96–99.

FRENCH, ELIZABETH G., & CHADWICK, IRENE. Some characteristics of affiliation motivation. *J. abnorm. soc. Psychol.*, 1956, **52**, 296–300.

HEYNS, R. W., ATKINSON, J. W., & VEROFF, J. A scoring manual for the affiliation motive. In J. W. Atkinson (Ed.), *Motives in fantasy, action and society*. Princeton: Van Nostrand, 1958.

KISH, L. Confidence intervals for clustered samples. *Am. sociol. Rev.*, 1957, **22**, 154–165.

SHIPLEY, T. E., & VEROFF, J. A projective measure of need for affiliation. *J. exp. Psychol.*, 1952, **43**, 349–356.

Received February 12, 1958.

RE-EVALUATION OF CHOICE ALTERNATIVES AS A FUNCTION OF THEIR NUMBER AND QUALITATIVE SIMILARITY[1]

JACK W. BREHM[2] AND ARTHUR R. COHEN

Yale University

Most theory and research relevant to choice situations have centered on prechoice phenomena. There is, however, increasing concern with postdecision phenomena. A theory recently proposed by Festinger (1957), for example, allows several specific derivations about the psychological consequences of choice. Some of these derivations have been tested (Brehm, 1956; Ehrlich, Guttman, Schonbach, & Mills, 1957; Festinger, 1957). This paper reports a test of two further derivations.

The central conception of this theory is that a person who holds cognitions about himself or the environment that are inconsistent with each other experiences psychological "dissonance," which he attempts to reduce or eliminate by trying to reduce the number of inconsistent cognitions in relation to the number of consistent ones. An approximate statement of the fundamental hypothesis as it applies to the choice situation is that *the magnitude of psychological dissonance is a direct function of what one has to give up compared to what one obtains.* That is, the attractiveness of an unchosen alternative is a cognition inconsistent with the knowledge that one has chosen a different alternative. In purchasing a new television receiver, for example, having to give up a new car would create more dissonance than having to give up a new pair of shoes.

This general statement allows two further derivations about the consequences of choice: (*a*) the greater the qualitative dissimilarity of alternatives, the greater the magnitude of dissonance from choice; and (*b*) the greater the number of alternatives, the greater the magnitude of dissonance from choice. Other things being equal, as qualitative dissimilarity increases, what one has to give up relative to what one gains increases. For example, choosing one of two automobiles of different makes (but of equal attractiveness) would create more dissonance than would choosing one of two that are identical except for color. More features such as type of styling, type of motor, etc., must be given up by a choice between different makes than by a choice between two nearly identical cars of the same make. The number of alternatives affects the magnitude of dissonance in much the same way. Where the alternatives have at least some qualitative dissimilarity, and with other factors held constant, the greater the number of alternatives from which one must choose, the more one must give up and consequently the greater the magnitude of dissonance.

When a person experiences psychological dissonance, he tries to reduce or eliminate it. If the dissonance has occurred as the consequence of making a choice, it may be reduced or eliminated by re-evaluating the alternatives so as to minimize what has been given up *relative* to what has been obtained. Thus to the extent to which an individual is experiencing dissonance, he may be expected to increase the perceived attractiveness of the chosen alternative and decrease the perceived attractiveness of unchosen alternatives.

The experiment described was designed to test the following two general hypotheses:

1. Other things being equal, the greater the qualitative dissimilarity of alternatives, the greater is the magnitude of dissonance created by choice, and the greater the consequent re-evaluation of these alternatives.

2. Other things being equal, the greater the number of alternatives, the greater is the magnitude of dissonance and the consequent re-evaluation of these alternatives.

METHOD

The strategy was to ask sixth-grade children to indicate, both before and after a choice, how much they liked each of several different toys. The choice given to about one half of the *Ss* was between similar toys, and to the other half it was between dissimilar toys. From each of these two groups about half of the *Ss* were given a choice between two toys, while the

[1] This is a report of a study conducted as part of the Yale Communications Research Program, and supported by grants to Carl I. Hovland from Bell Telephone Laboratories and from the Rockefeller Foundation.

[2] Now at Duke University.

other half were given a choice of one among four toys. *Ss*' average change in liking of the chosen and unchosen alternatives was then computed for each kind of choice.

Subjects. The original sample consisted of 96 boys and 107 girls in eight different classes in three different schools,[5] who were present for both the first and second sessions of the experiment. A selection procedure, to be described later, reduced the usable number to 29 boys and 43 girls.

Procedure. The experimenter (*E*) and his helper were introduced to each class as persons from Yale University who were doing a survey having to do with toys. *E* then said to the *Ss*:

> Some toy manufacturers have recently asked us to find out what kinds of toys people like you like and what kinds you don't like. Christmas is not very far away and so it's especially important for the toy manufacturers to know what you like or dislike so they will know what to make and send to all the stores.
>
> Now in order to find what you really do like or dislike we are going to ask each of you to indicate on a questionnaire just how much you like or dislike each of 16 different toys.

After the questionnaires had been distributed and explained, the 16 toys, all in one large box, were brought into the room. Each toy had a number to correspond with the number on one of the rating scales on the questionnaire. *E* then went through the toys by numbered order, holding each long enough and close enough so that it could be inspected closely by every child. At the beginning of this procedure *Ss* were told that their own personal answers were wanted, not what their neighbor thought, so they were asked not to talk while filling out the questionnaire. One or two gentle reminders were usually necessary during the course of the rating. When the ratings were completed, the questionnaires were collected and the *Ss* thanked for helping with the survey. No mention was made of the possibility that *E* and his helper would return.

The second session took place exactly one week after the first. It was conducted by the same *E*, though for some classes a different person had to be used as the helper.[6] *E* first reminded the *Ss* that he and his helper were from Yale and were conducting a survey for some toy manufacturers. He then went on to say:

> The information you gave us last week was very good. In fact, it was so good that the toy manufacturers have asked us to get some more information. So today we're going to ask you to help us out by answering some more questions about the toys.
>
> Now since the toy manufacturers were so pleased with what you did for them last week, and since they are asking you to do something more for them this week, they thought it only fair that they give you something for all this help. So they decided to give each one of you one of the toys that you've been rating. Unfortunately, however, we don't have enough of any one toy so that you could all have the same thing. That means we can't let you choose just anything you want. So what we did was to pick two (or four) for each of you to choose between. While we've tried to give you a choice between things you like, you probably won't get a chance to choose the one you like most. We just couldn't arrange the choices that well.

Ss were then given prearranged choice sheets that indicated the toys, by name and number, from which they could choose. They were cautioned not to talk about their choices. They were then shown each of the 16 toys, one by one, and asked to circle which one of the two (or four) on their choice sheet they wanted as their reward for helping us. The toys, they were told, would be delivered to the school the same week.

After the choices had been made and the choice sheets collected, *E* explained what additional help the *Ss* were being asked to give. To do this he said:

> You'll remember I said we were going to ask you to do something more for us this week. Well, what we want you to do is to indicate for us again just how much you like or dislike each of these 16 toys. You see a person's liking for toys like these may change from one week to the next. You might like a toy more, less, or about the same as you did last week and it's important for the toy manufacturers to know about this. So we are now going to ask each one of you to fill out a questionnaire just like the one you filled out last week.[7]

The toys were again held up, one by one, but this time somewhat more rapidly since they were already familiar to the *Ss*. Upon completion of the ratings the questionnaires were collected.

At this, the end of the experiment, the deception, design, and purpose were completely explained to the *Ss*. It was further explained that the experimenters were unable to afford a toy for each person but that instead the toys would be divided up among the eight classes participating, so that each class would receive four toys for use in the class room itself.[8]

Manipulation of Number of Alternatives. As indicated in the above instructions, the *Ss* in some classes were told they could choose between two of the toys, and their individual choice sheets designated the two

---

[5] The authors wish to express their great appreciation to the following people for making this study possible: Charles T. St. Clair, Superintendent of North Haven Schools, Thomas P. Ryan, Principal at Temple School, Mildred A. Wakeley, Principal at Center School, and Edward L. White, Principal at Ridge Road School, and all of the sixth-grade teachers at these schools.

[6] The authors would like to thank Mary Lee Brehm and Robert Ascheim for acting as helper at some of the experimental sessions.

[7] The first class run was additionally asked to write out as much as they could remember of the characteristics of the choice objects. However, the *Ss* were unable to handle this task so no further groups were asked to do it.

[8] While the school officials and teachers involved in this study reported no apparent ill effects, the authors were somewhat dissatisfied with their ability to reduce the *S*'s residual tension. We now believe that a major deception of this type should be used at this age level only if the catharsis can be conducted individually or in small groups so the *E* can make sure of its success.

toys by name and number. *Ss* in other classes were told they could choose one from four toys that again were designated on their choice sheets by name and number.

Manipulation of Qualitative Similarity of Alternatives. The 16 toys rated and used as choice alternatives were chosen a priori to be relatively similar or dissimilar to each other. Two sets of four toys were picked so that there would be high similarity among the four in each set. The remaining eight toys were picked to be relatively dissimilar to any in the two sets of four, and to most of the other seven. The criterion of qualitative similarity was the similarity of activity that the toy would require, as indicated on the following classified list in which the toys appear in the same order as in the experiment:

1. Swimming Fins
2. Swimming Mask
3. Snorkel
4. Life Jacket (All for water activities)
5. Set of 52 Games
6. Set of 66 Games
7. Set of 52 Games (slightly different from #5)
8. Set of 58 Games (All indoor table games)
9. Indoor Bowling
10. Ship Model
11. Copper Craft
12. Racing Game
13. Aluminum Craft
14. Erector Set
15. Archery Set
16. Metal Tapping Craft

To create relative similarity, all of the alternatives offered were chosen from Numbers 1–4, or from 5–8. To create dissimilarity, all of the alternatives were chosen so they involved different activities, e.g., Numbers 1, 5, 14, and 16.

Questionnaires. The main rating questionnaire consisted of 16 rating scales, each numbered for one of the toys. Each scale consisted of a horizontal line of 51 equally spaced dots with six equally spaced markers labeled: Don't like at all; Like just a little; Like fairly well; Like quite a lot; Like very much; Really like very, very much. Numerical values were given to the ratings by assigning values 0 through 5.0 to these labels, respectively, and scoring each answer to the nearest tenth of an interval (i.e., to the nearest dot on the scale). *Ss* were instructed to make a pencil check mark anywhere along the dotted line. There was a space for name and sex at the top of each questionnaire.

Design. Half of the eight classes were run in the Two Alternative condition, the other half, in the Four Alternative condition. Half of the *Ss* in each class were given Similar Alternatives, the other half, Dissimilar Alternatives.

The classes at any one school were run during a single morning on the same week day but one week apart for the two experimental sessions. Work at the three schools was conducted on three successive days.

Controls. While all *Ss* were given a choice and asked to fill out the questionnaires, only a portion of them yielded usable data. Many were excluded by requirements that had to be met to permit an appropriate test of the hypotheses. First, the choice alternatives for each *S* had to be liked, but not so much that an increase in liking would be impossible. Second, one alternative had to be initially more liked than any other so that its choice could be expected. Increased ratings of the chosen item are thus not likely to be simply a result of normal (and random) changes in *actual* liking from the first questionnaire to the second. In addition, *Ss* who failed to choose the alternative initially marked as most liked were excluded because they gave unreliable or invalid ratings. Finally, in order to ensure that initially less liked alternatives were seriously considered as possible choices, initial ratings of these alternatives could not be much lower than the most liked alternative.

The stringency with which these requirements were applied was determined by the actual distribution of initial ratings and the need for a workable number of *Ss* in each of the four experimental conditions. The final criteria were: (*a*) the most liked alternative had to receive a rating between 2.9 and 4.6, or in other words, a moderate to high Liking score; (*b*) the next most attractive alternative could be no more than 1.1 scale points lower, but had to be at least 0.1 scale points (the least measureable difference) lower, than the first; (*c*) for four alternatives, the additional criterion was imposed that the least liked must be no more than 2.5 scale points below the most liked alternative.

TABLE 1

CHANGES IN LIKING OF CHOSEN AND INITIALLY NEXT MOST LIKED ALTERNATIVE

Number of Alternatives	Qualitative Character			
	Similar	*N*	Dissimilar	*N*
Two				
Chosen	0.07[a]	29	0.20	21
Unchosen	−0.09	29	−0.65	21
Four				
Chosen	0.48	10	0.62	12
Unchosen	−0.27	10	−0.72	12

[a] A positive number indicates increased Liking, a negative, decreased Liking.

RESULTS

Liking change scores were computed for all choice alternatives for the selected *Ss*. In order to compare the Four Alternative conditions with the Two, we have taken the chosen and the initially next most attractive alternative of the remaining three. Our hypotheses predict increased liking for the chosen alternative and decreased liking for the unchosen alternative, to the extent that dissonance is created by the choice. Thus we would expect greater change in liking in the Four than in the Two Alternative condition, and greater change in liking in the Dissimilar than in the Similar Alternative condition.

Table 1 presents the relevant data. Both hypotheses are supported. The chosen alternative in each of the four experimental conditions shows an increase in Liking, while the unchosen

TABLE 2
TOTAL DISSONANCE REDUCTION (CHANGES IN LIKING) FOR COMBINED EXPERIMENTAL CONDITIONS

	Dissonance Reduction	*N*
Number of Alternatives		
Two	0.45	50
Four	1.08	22
Qualitative Character		
Similar	0.31	39
Dissimilar	1.03	33

alternative always shows a decrease in Liking. If we take the sum of changes in Liking for chosen and unchosen alternative as an index of total reduction in dissonance and combine conditions in order to arrive at over-all dissonance reduction scores for each of the two experimental manipulations, then we obtain the figures given in Table 2. Here it is clear that the Four Alternative condition produced greater dissonance reduction than did Two (p equals about .02),[7] and the Dissimilar Alternative condition produced more dissonance reduction than did the Similar (p less than .01).[8]

The criteria used in selecting alternatives and *S*s produced a small bias in the initial difference between the two alternatives in question such that the initial difference in the Four Alternative condition was less than that in the Two. Since the initial relative attractiveness of alternatives has been shown to determine amount of dissonance reduction (Ehrlich, Guttman, Schonbach, & Mills, 1957) it is necessary to investigate whether or not the present bias can alternatively account for the obtained results. When this is done by selecting out of the Four Alternative conditions those *S*s for whom the initial difference was relatively large and comparing their data with those for *S*s whose initial difference was small, the mean score for total dissonance reduction turns out to be the same regardless of size of initial difference. This small bias thus does not account for the obtained difference between the Two and Four Alternative conditions.

Changes in Liking for the Third and Fourth Alternatives. In the Four Alternative conditions, it is possible to further investigate the effects of Similar versus Dissimilar alternatives by comparing changes in Liking of the third and fourth alternatives. One might expect that such changes for these somewhat less attractive alternatives would follow the same pattern that held for changes in Liking of the previously considered unchosen alternative, but with smaller magnitude because of their lesser importance.

Mean change in Liking was therefore computed for the third and fourth alternatives in the Similar and Dissimilar conditions. However, since the initial rating of these alternatives tends to be below average, some positive change in rating may be expected solely from regression effects. In order to obtain a better estimate of Liking change it is necessary to correct the rating change by subtracting the estimated regression. Estimates of regression were obtained by computing the mean Liking change according to amount of initial rating for all toys not involved in the choice.[9]

The corrected Liking changes for the third and fourth alternatives, presented in Table 3, show an over-all increase in liking rather than the expected decrease. Furthermore, there is no consistent difference between the Similar and Dissimilar conditions. Our expectations for the third and fourth alternatives are clearly disconfirmed by these results.

[7] All p values reported were obtained by two-tail t tests.

[8] The same computations for our unselected population yield support for the effect of the Similar–Dissimilar manipulation, but fail to show any effect of the number of alternatives. The very reason for selection of *S*s was, of course, that data from these unselected *S*s would fail to adequately test the hypotheses. Consequently, no conclusion should be drawn from the data for these unselected *S*s.

[9] To relate rating change to the initial rating, the Liking scale was divided into the following intervals: 0–0.5; 0.6–1.5; 1.6–2.5; 2.6–3.5; 3.6–4.5; 4.6–5.0. The mean Liking changes then obtained from all selected *S*s for toys not involved in the choice and with initial ratings in each of the above categories were: +.5; +.5; +.1; −.1; −.3; and −.5, respectively. These corrections were then subtracted from each individual's Liking change scores according to his initial ratings of each of the alternatives.

Use of these corrections for the chosen and most-liked unchosen alternatives reduces slightly the obtained p values but in no way alters the pattern of dissonance reduction scores or conclusions that may be drawn from them. Thus with these corrections also, greater liking changes occur for dissimilar than similar and for four than two alternatives.

TABLE 3
CORRECTED LIKING CHANGES FOR THE THIRD AND FOURTH ALTERNATIVES

Qualitative Character	Alternative		N
	Three	Four	
Similar	0.10	0.94	10
Dissimilar	0.58	0.21	12

DISCUSSION

The primary data from the present experiment support two derivations from Festinger's theory of cognitive dissonance: the magnitude of dissonance and consequent re-evaluation of choice alternatives is a direct function of (*a*) the number of choice alternatives and (*b*) the qualitative dissimilarity of the choice alternatives.

The effect of our manipulation of the number of alternatives may, however, be explained in a different way. It will be recalled that the magnitude of dissonance and consequent re-evaluation of choice alternatives is a function of the difference in initial liking for the alternatives: the smaller this initial difference, the greater the magnitude of dissonance after the choice. It follows that with other things held constant, the more liked are the unchosen alternatives, the greater the magnitude of dissonance. It is clear, then, that the Two and Four Alternative conditions must be equated on the initial liking of both the chosen and most liked unchosen alternatives. This was done, of course, as part of our procedure for selecting *S*s whose choices would furnish a good test of the hypotheses. But the low reliability of our measure of liking means that where there are three unchosen alternatives, the one that was marked as initially most liked will not always be the one that was *actually* most liked. Thus the *actual* liking for the most liked unchosen alternative should tend to be higher where there are three unchosen alternatives than where there is only one. In addition, let us assume that, with liking for the magnitude of the chosen alternative held constant, the most liked unchosen alternative is a disproportionately more important determinant of dissonance than are the less liked alternatives. If this assumption is correct, then the difference in liking for the unchosen alternative implies that dissonance should have been greater in the Four than in the Two Alternative condition. It seems unlikely, however, that this bias could have been large enough to have produced the observed effects. While this alternative explanation is possible, it is not entirely plausible.

A more serious question may be raised in regard to the data for the third and fourth choice alternatives. For while the re-evaluation of the chosen alternative clearly reflects greater dissonance in the Four than in the Two Alternative condition (p equals .04), the third and fourth alternatives fail to show any sign of dissonance reduction. Indeed, the increase in liking for these alternatives should lead to *increased* dissonance. One possible explanation for this apparent increase is that our estimate of regression is too small. It may be, for example, that when a toy is made an alternative for choice, it tends, prior to the choice, to be re-evaluated more than a toy not made one of the alternatives. If the direction of such re-evaluations was random, our estimate of regression would be too small, thus accounting for the apparent increase in liking. It is also possible that such re-evaluation is not random, but rather is directed toward enhancing the value of all alternatives. The latter explanation would not only account for the increased liking of the third and fourth alternatives; it would also account for the fact that the chosen alternative increases in liking more than the initially most liked unchosen alternative decreases. While the phenomenon can thus be plausibly explained, it is of sufficient importance to the theory of dissonance to warrant separate and thorough investigation.

It is also of interest that the present experiment produced no visible interaction of the two independent variables in their determination of the magnitude of dissonance created by choice. Under certain conditions such an interaction would be expected. Thus if the qualitative similarity of alternatives is extreme (approaching identity) little or no dissonance would be created by choice no matter how many alternatives were involved. But from that point, increasing the dissimilarity should have more effect on dissonance where several alternatives are involved than where only two

are involved. Though we have no independent evidence from the present experiment, we may assume that the alternatives in the Similar condition were seen as being somewhat dissimilar. If this is the reason that no interaction effect was obtained, then a good test of the interaction hypothesis must employ a much more powerful manipulation of qualitative similarity than was used here.

SUMMARY

In order to test two derivations from Festinger's theory of psychological dissonance, sixth-grade school children were given systematically varied choices from among several attractive toys. About half of them were given a choice between two toys, and the remainder were given a choice of one from among four toys. From each of these two groups, about half was given a choice between qualitatively similar toys, while the remainder was given a choice between qualitatively dissimilar toys. *S*s were asked, both before and after the choice, to indicate on a 51-point graphic rating scale how much they liked each of the toys.

The two derivations tested were: the magnitude of dissonance and consequent amount of increased liking for the chosen alternative and decreased liking for the unchosen alternative increases with (*a*) the qualitative dissimilarity of alternatives, and (*b*) the number of alternatives. The first of these was confirmed; the second received somewhat questionable support. Changes in liking for the third and fourth unchosen alternatives failed to support our expectations. Two explanations for this failure were suggested.

REFERENCES

BREHM, J. W. Postdecision changes in the desirability of alternatives. *J. abnorm. soc. Psychol.*, 1956, **52**, 384–389.

EHRLICH, DANUTA, GUTTMAN, I., SCHONBACH, P., & MILLS, J. Postdecision exposure to relevant information. *J. abnorm. soc. Psychol.*, 1957, **54**, 98–102.

FESTINGER, L. *Theory of cognitive dissonance.* Evanston: Row, Peterson, 1957.

Received March 17, 1958

INCREASING COGNITIVE DISSONANCE BY A *FAIT ACCOMPLI*[1]

JACK W. BREHM[2]

Yale University

A RECENT theory by Festinger (1957) attempts to state the conditions under which discrepancies between cognitions produce "cognitive dissonance," a state of tension that motivates the individual to reduce or eliminate the discrepancies. One of the conditions that may be necessary for the creation of cognitive dissonance is that there be some element of volition. The theory states, for example, that a person who is completely forced to behave in a manner he would avoid if possible, experiences no dissonance. On the other hand, a *fait accompli*—i.e., an event outside of the person's control—might conceivably create dissonance if that same event would have led to the opposite behavior had it been predictable at a prior choice point. The purpose of this paper, then, is to report an experiment in which a *fait accompli* does appear to have increased cognitive dissonance.

The theoretical expectations may be stated succinctly. If a person is induced by the promise of a small prize to eat a disliked food, he should experience dissonance. If other things are held constant, the amount of dissonance can be increased by increasing the amount of the disliked food to be eaten. A person who is experiencing such dissonance should try, in proportion to its magnitude, to reduce it by convincing himself that he likes the food more than he originally thought.

If, then, the amount of expected eating is increased by a *fait accompli* after the initial decision to eat has occurred, and this is accompanied by increased liking, it may be concluded that a *fait accompli* can affect the magnitude of dissonance.

METHOD

In order to examine the effect of a *fait accompli*, eighth-grade students were offered attractive prizes for eating a small portion of a disliked vegetable. About half of them were told, after eating the vegetable, that their parents would be informed which vegetable they ate. Liking for the vegetable was measured before and after this procedure.

The pre-experimental questionnaire. Approximately three weeks prior to the beginning of the experiment, questionnaires were given by teachers to all eighth-grade students in a New Haven Junior High School.[3] Students were told that the questionnaire was sponsored by a nonprofit organization interested in people's eating habits. The questionnaire asked for each of 34 vegetables, "Considering everything about it (taste, smell, appearance, texture), to what extent do you like or dislike this vegetable?" Answers were indicated by making an X on any one of 61 dots running along a straight line and identified each 10 points by the following labels: like very much; like a lot; like a little; neither; dislike a little; dislike a lot; dislike very much. Subjects' responses were then scored from plus 3.0 for "like very much" to minus 3.0 for "dislike very much."

Experimental procedure. Subjects (*S*s) were given passes during school hours to come one at a time to take part in a food-testing program. When an *S* arrived, the experimenter (*E*) introduced himself as being a representative of a fictitious consumer organization, whose business it was to test products and publish reports on them in order to help people in their shopping. It was explained that the organization was at present conducting a study of a number of vegetables by testing each for its food value in terms of vitamins and minerals and also by asking all kinds of people what they thought about them. *E* then said that what he wanted the *S* to do was answer some questions but only about one vegetable in order to get good information about that one. Before being given the main experimental questionnaire, the *S* was asked to answer some simple questions about his age, his favorite and most disliked foods, how often the vegetable picked for him was served in his home, and how often he ate it when it was served. It was also explained at this point that the vegetable in question had been picked at random, though actually it had been picked because it was one of the ones he most disliked (it was at least at −2.0 on the prequestionnaire). The *S* was then asked to fill out the main questionnaire carefully and accurately. To stress the anonymous nature of the questionnaires, the *S* was given a plain Manila envelope in which to put them when completed.

After the main questionnaire was completed, *E* said,

> While we feel we get good information by asking a person to carefully fill out a questionnaire this way, we also know that we may get different kinds of

[1] This study was conducted as part of the Yale Communication Research Program and was supported by a grant to Carl I. Hovland from Bell Telephone Laboratories. The author would like to thank Arthur R. Cohen, Carl I. Hovland, Leon Festinger, and Irving Sarnoff for their comments on the manuscript.

[2] Now at Duke University.

[3] For their help in conducting this study, the author wishes to thank Louisa Willard, Psychologist for the New Haven Schools, Frank Carr, Principal of Troupe Junior High School, Mrs. Connolly, and the eighth-grade teachers, especially Mrs. Himmelfarb.

are involved. Though we have no independent evidence from the present experiment, we may assume that the alternatives in the Similar condition were seen as being somewhat dissimilar. If this is the reason that no interaction effect was obtained, then a good test of the interaction hypothesis must employ a much more powerful manipulation of qualitative similarity than was used here.

SUMMARY

In order to test two derivations from Festinger's theory of psychological dissonance, sixth-grade school children were given systematically varied choices from among several attractive toys. About half of them were given a choice between two toys, and the remainder were given a choice of one from among four toys. From each of these two groups, about half was given a choice between qualitatively similar toys, while the remainder was given a choice between qualitatively dissimilar toys. *Ss* were asked, both before and after the choice, to indicate on a 51-point graphic rating scale how much they liked each of the toys.

The two derivations tested were: the magnitude of dissonance and consequent amount of increased liking for the chosen alternative and decreased liking for the unchosen alternative increases with (*a*) the qualitative dissimilarity of alternatives, and (*b*) the number of alternatives. The first of these was confirmed; the second received somewhat questionable support. Changes in liking for the third and fourth unchosen alternatives failed to support our expectations. Two explanations for this failure were suggested.

REFERENCES

BREHM, J. W. Postdecision changes in the desirability of alternatives. *J. abnorm. soc. Psychol.*, 1956, **52**, 384–389.

EHRLICH, DANUTA, GUTTMAN, I., SCHONBACH, P., & MILLS, J. Postdecision exposure to relevant information. *J. abnorm. soc. Psychol.*, 1957, **54**, 98–102.

FESTINGER, L. *Theory of cognitive dissonance.* Evanston: Row, Peterson, 1957.

Received March 17, 1958.

INCREASING COGNITIVE DISSONANCE BY A *FAIT ACCOMPLI*[1]

JACK W. BREHM[2]

Yale University

A RECENT theory by Festinger (1957) attempts to state the conditions under which discrepancies between cognitions produce "cognitive dissonance," a state of tension that motivates the individual to reduce or eliminate the discrepancies. One of the conditions that may be necessary for the creation of cognitive dissonance is that there be some element of volition. The theory states, for example, that a person who is completely forced to behave in a manner he would avoid if possible, experiences no dissonance. On the other hand, a *fait accompli*—i.e., an event outside of the person's control—might conceivably create dissonance if that same event would have led to the opposite behavior had it been predictable at a prior choice point. The purpose of this paper, then, is to report an experiment in which a *fait accompli* does appear to have increased cognitive dissonance.

The theoretical expectations may be stated succinctly. If a person is induced by the promise of a small prize to eat a disliked food, he should experience dissonance. If other things are held constant, the amount of dissonance can be increased by increasing the amount of the disliked food to be eaten. A person who is experiencing such dissonance should try, in proportion to its magnitude, to reduce it by convincing himself that he likes the food more than he originally thought.

If, then, the amount of expected eating is increased by a *fait accompli* after the initial decision to eat has occurred, and this is accompanied by increased liking, it may be concluded that a *fait accompli* can affect the magnitude of dissonance.

METHOD

In order to examine the effect of a *fait accompli*, eighth-grade students were offered attractive prizes for eating a small portion of a disliked vegetable. About half of them were told, after eating the vegetable, that their parents would be informed which vegetable they ate. Liking for the vegetable was measured before and after this procedure.

The pre-experimental questionnaire. Approximately three weeks prior to the beginning of the experiment, questionnaires were given by teachers to all eighth-grade students in a New Haven Junior High School.[3] Students were told that the questionnaire was sponsored by a nonprofit organization interested in people's eating habits. The questionnaire asked for each of 34 vegetables, "Considering everything about it (taste, smell, appearance, texture), to what extent do you like or dislike this vegetable?" Answers were indicated by making an X on any one of 61 dots running along a straight line and identified each 10 points by the following labels: like very much; like a lot; like a little; neither; dislike a little; dislike a lot; dislike very much. Subjects' responses were then scored from plus 3.0 for "like very much" to minus 3.0 for "dislike very much."

Experimental procedure. Subjects (*S*s) were given passes during school hours to come one at a time to take part in a food-testing program. When an *S* arrived, the experimenter (*E*) introduced himself as being a representative of a fictitious consumer organization, whose business it was to test products and publish reports on them in order to help people in their shopping. It was explained that the organization was at present conducting a study of a number of vegetables by testing each for its food value in terms of vitamins and minerals and also by asking all kinds of people what they thought about them. *E* then said that what he wanted the *S* to do was answer some questions but only about one vegetable in order to get good information about that one. Before being given the main experimental questionnaire, the *S* was asked to answer some simple questions about his age, his favorite and most disliked foods, how often the vegetable picked for him was served in his home, and how often he ate it when it was served. It was also explained at this point that the vegetable in question had been picked at random, though actually it had been picked because it was one of the ones he most disliked (it was at least at −2.0 on the prequestionnaire). The *S* was then asked to fill out the main questionnaire carefully and accurately. To stress the anonymous nature of the questionnaires, the *S* was given a plain Manila envelope in which to put them when completed.

After the main questionnaire was completed, *E* said,

> While we feel we get good information by asking a person to carefully fill out a questionnaire this way, we also know that we may get different kinds of

[1] This study was conducted as part of the Yale Communication Research Program and was supported by a grant to Carl I. Hovland from Bell Telephone Laboratories. The author would like to thank Arthur R. Cohen, Carl I. Hovland, Leon Festinger, and Irving Sarnoff for their comments on the manuscript.

[2] Now at Duke University.

[3] For their help in conducting this study, the author wishes to thank Louisa Willard, Psychologist for the New Haven Schools, Frank Carr, Principal of Troupe Junior High School, Mrs. Connolly, and the eighth-grade teachers, especially Mrs. Himmelfarb.

answers to these questions if a person actually tastes some of the vegetable before filling out the questionnaire. So we would like to have everyone eat a little of the vegetable and then answer some further questions about it. But we also realize that you might not like to eat just anything we put in front of you so the rest of this is completely up to you. You can stop right here without doing anything more at all, or you can eat a small dish of the vegetable and answer some more questions. However, to get as many people as possible to give us this further information, we are offering a prize to those who do go on. The prize consists of either two 45 or 78 rpm records of your own choice, or two tickets for movies of your own choice. You can do whichever you want. You can leave right now, or you can eat a small dish of —— (name of vegetable) and get whichever of the two prizes you want, two phonograph records or two movie tickets.

If the *S* chose to quit he was thanked for his help and allowed to return to class. If he chose to continue, he was served a small dish of the vegetable, heated if appropriate. This constituted the *Low Consequence* experimental condition.

In the *High Consequence* condition, *S*s were given exactly the same instructions. However, in order to create the impression that their choice to eat the vegetable would result in consequent further eating of it, they were casually given, when nearly finished eating, these further instructions:

> Oh, I almost forgot to mention that one of the reports we plan to put out from this study will simply be a letter to the parents of each person who takes part, just indicating which vegetable that person ate.[4]

This information, which would likely have made many hesitate to go ahead with the eating if they had known about it, may be considered a *fait accompli*.

When the *S* had finished eating, he was asked to fill out again the main questionnaire. In making the request *E* mentioned that some people changed one way, some another, and some didn't change at all. He also stressed that the *S* should be as careful and accurate as possible and, finally, that the second questionnaire, like the first, was completely anonymous and to be slipped into the Manila envelope when completed.

When the *S* had finished the questionnaire, he was asked to choose which prize he wanted and told that the prizes would be sent in two or three weeks. He was also asked not to talk to other students about what happened in the testing procedure.[5]

When the study was completed, its true purpose was fully explained to all *S*s. The prizes were given as promised.

The experimental questionnaires. The first experimental questionnaire contained two important questions. The first of these asked *S*s to indicate on a 51-point line with six equally spaced labels running from "every meal" to "never," how often the vegetable chosen for them was served at home. The second question, using an identical scale, asked how often they ate the vegetable when it was served. The main experimental questionnaire contained a question identical to that used on the pre-experimental questionnaire to measure liking for the vegetable.[6]

Design. In all, 20 *S*s were run in the Low Consequence condition, 19 were run in the High Consequence condition. While an attempt was made to use the vegetables with equal frequency, this was only partly successful. Inspection of the dependent variables, however, reveals no large differences between vegetables.

RESULTS

Eating a small amount of a disliked vegetable should create some dissonance and consequent increase in liking in order to reduce the dissonance. If, in addition, the *fait accompli* (the High Consequence manipulation) increases dissonance, then persons exposed to it should experience greater dissonance and thus increase their liking for the vegetable more than those who are not exposed to it.

The mean change in Liking was therefore computed for *S*s in each of the two experimental conditions. The prediction was confirmed in that the mean increased Liking for the High Consequence condition (3.76) was greater than the mean increased Liking for the Low Consequence condition (2.32). The difference between the conditions is significant at beyond the .01 level.[7]

While the instructions for the High Consequence condition do not obviously advocate liking the vegetable, it is conceivable they were interpreted in that way by the *S*s. The differential increased liking for the vegetable between the two conditions might simply reflect differential "endorsement" of the vegetable by the "consumer organization." Thus our original explanation in terms of "dissonance reduction" requires further support to be cogent. We must therefore examine the manipulation and its effect more closely.

---

[4] The idea for this manipulation came from some pretest *S*s who spontaneously said, "Boy, would my mother like to know I'm eating this!"

[5] Informal interviewing of *S*s in the later part of the experiment indicated that there was a moderate amount of talking about the experiment. Most *S*s had heard, before participating, what some of the vegetables were, that some people actually ate the vegetables, that prizes were being given away, and what the prizes were. None showed evidence, however, of being aware of the High Consequence Manipulation.

[6] The main questionnaire also contained several questions included to gather data for a larger study of which this experiment was a part.

[7] All p values in this paper were obtained by two-tailed t tests.

It will be remembered that right at the beginning of the experiment all *S*s were asked to show how often their vegetable was served at home, and how often they ate it when it was served. If the "dissonance" explanation is correct, then the High Consequence manipulation, which depends on an implied increase in frequency of eating at home, should work best for *S*s who indicated a large discrepancy between frequency of serving and frequency of eating. *S*s who indicated little or no discrepancy between serving and eating should have experienced little or no increase in dissonance from the High Consequence manipulation. Therefore, those *S*s who indicated a high discrepancy in frequency should show a greater increase in Liking from the Low to the High Consequence condition than should those who indicated a low discrepancy in frequency. If, on the other hand, the "endorsement" explanation is correct, there would be no reason to expect differential results between *S*s indicating high, and *S*s indicating low, discrepancy.

To make the indicated analysis, *S*s within each experimental condition were split at the median into high and low discrepancy groups. For each of these groups the mean change in Liking was computed. It was found that for those individuals who indicated a low discrepancy in frequencies, the mean difference in Liking change (.67) between Low and High Consequence conditions was negligible, while for those who indicated a high discrepancy the difference (2.01) was significant at beyond the .01 level. These data thus support the conclusion that the High Consequence manipulation implies an increased frequency of eating at home and thereby increases the magnitude of dissonance and pressure to increase one's liking.

Further examination of the present data for evidence relevant to the "dissonance" explanation turned up an interesting phenomenon. *S*s within each condition were divided, as close to the median as possible, according to their initial dislike for their vegetable. For each of these subgroups the final mean Liking score was computed.[8] In this way it was found that within both conditions those persons who indicated the greater initial dislike subsequently indicated greater Liking than did those with less initial dislike. Within the High Consequence condition, this difference is significant beyond the .01 level, and within the Low Consequence condition the difference yields a two-tailed t of about 1.0. It is improbable that persons with extreme initial dislike have less resistance to increasing their liking than do those with moderate initial dislike. We may therefore conclude that the greater was the individual's initial dislike, the greater was the experimentally produced pressure on him to increase his liking for the vegetable. This finding is consistent with dissonance theory, for as initial dislike increases, so should the dissonance from having to eat the vegetable.

[8] Increased Liking from regression should increase with more extreme initial dislike and thus make the *change* scores unequal. For this reason, *final* Liking scores, which are not biased by regression or ceiling effects, were used for this analysis.

DISCUSSION

Other research on the effects of a *fait accompli* has been imprecise as a result of the limitations of field studies (e.g., Clark, 1953), or, if done in the laboratory, has failed to produce experimentally adequate information (e.g., Brehm & Cohen, 1959). The present experiment, however, unequivocally demonstrates attitude change from a *fait accompli*, which is relevant to prior induction of unpleasant behavior.

If the differential increases in liking in the present experiment were indeed a result of differential amounts of dissonance, then the conceptual definition of dissonance has been further specified. The specification is not complete, however, for it may be that a *fait accompli* affects the magnitude of dissonance only when it is related to a previous choice. Further research is necessary to supply the answer to this and similar hunches.

A finding of perhaps equal theoretical interest was the relationship between initial strength of attitude and subsequent change. The more unfavorable was the initial attitude, the more favorable it tended to be at the end of the experiment. This relationship would be inconsistent with any theory of attitude change that ignores the size of discrepancy between positions held and advocated. It is, however, entirely consistent with dissonance theory. As the discrepancy between attitude and behavior increases, the amount of dissonance

and consequent pressure to reduce it increase. Thus, the greater is the discrepancy between attitude and behavior, the greater will be the subsequent pressure to change the attitude. There is, of course, some resistance to changing one's attitude. The magnitude of this resistance might be expected to increase as the individual's position becomes more extreme. Under the conditions of the present experiment, the pressure to reduce dissonance apparently increased more rapidly with increasing discrepancy than did the resistance against change. This experiment centered around a relatively individualistic attitude (food preference). If, on the other hand, the initial attitude were supported by stronger, external sources of reality, it might not be overcome by the pressure to reduce dissonance. Again, further research is needed to explore the problem.

SUMMARY

This paper presents a study illustrating the effect of a *fait accompli* on cognitive dissonance. Eighth-grade boys and girls were induced by the promise of a prize to eat disliked vegetables. Some were additionally told, while eating, that their parents would learn what vegetable they had eaten, with the implication the *S*s would have to eat more at home. A questionnaire measured liking for the vegetable before and after the experiment.

It was found that the *fait accompli* of implied further eating produced a greater increase in liking. This experiment therefore further specifies Festinger's concept of cognitive dissonance. In addition, it was found that persons initially least favorable toward the vegetable tended to be most favorable after the experiment. This result was seen as consistent with dissonance theory but not with some other theories of attitude change.

REFERENCES

BREHM, J. W., & COHEN, A. R. Choice and chance relative deprivation as determinants of cognitive dissonance. *J. abnorm. soc. Psychol.*, 1959, **58**, 383–387.

CLARK, K. B. Desegregation: An appraisal of the evidence. *J. soc. Issues*, 1953, **9**, 43–79.

FESTINGER, L. *A theory of cognitive dissonance.* Evanston: Row, Peterson, 1957.

Received March 19, 1958.

CHOICE AND CHANCE RELATIVE DEPRIVATION AS DETERMINANTS OF COGNITIVE DISSONANCE[1]

JACK W. BREHM[2] AND ARTHUR R. COHEN

Yale University

IT IS not clear from Festinger's (1957) theory of cognitive dissonance whether events capable of producing dissonance under conditions of choice might fail to produce dissonance if the person could not exercise some choice in controlling its occurrence or his exposure to it. It has been shown, for example, that to the extent that a person is *forced* to do something he ordinarily would avoid, he experiences no dissonance (e.g., Cohen, Brehm, & Fleming, 1959). It would seem, then, that a chance event that is outside the person's control may fall in the class of phenomena that fail to produce dissonance. Similarly, a *fait accompli* is outside his control and may produce no dissonance. An experiment by Brehm (1959), however, demonstrates that a *fait accompli* can affect the magnitude of dissonance created by a prior choice. Thus, while the limits of conditions that create dissonance may be greater than those stipulated by Festinger, just where these limits lie is not yet known. The purpose of the present paper is to report an experiment designed to establish some of these limits.

Two questions will be examined in the present study: Can a chance event affect the magnitude of dissonance? Does the effect of such a chance event depend upon there having been a prior relevant choice? In order to clarify these issues, we must discuss two theoretical approaches to the setting in which the test was made.

The first is Merton's analysis (Merton & Kitt, 1950) of the relationship between "relative deprivation" and felt satisfaction. According to this view, the degree to which a person is satisfied with a given state of affairs may be determined not necessarily by his absolute level of achievement or benefit, but instead by how much he achieves or benefits *relative to those around him*. The greater is the relative deprivation, the less will be the felt satisfaction with any arbitrary assignment or *fait accompli*.

The second theoretical approach is Festinger's analysis (1957; Festinger & Carlsmith, 1959) of forced compliance situations. This approach asserts that a person who is induced to behave in a way he ordinarily would avoid, experiences "dissonance"—i.e., psychological tension—which he then tries to reduce or eliminate. Other things being equal, the more unpleasant the induced behavior, the greater the magnitude of dissonance. Thus as relative deprivation increases, the magnitude of dissonance should increase. If a person is induced to perform an unpleasant behavior, thus creating dissonance, he can reduce or eliminate the dissonance by convincing himself that the behavior is really quite satisfactory after all (Brehm, 1957). His final attitude toward the event is therefore in part a direct function of the amount of dissonance created. The greater the relative deprivation, the greater is the amount of dissonance created and, consequently, the more positive his attitude toward the event.

With a setting of relative deprivation, then, we can test the points of theory at issue. By varying the amount of force used to make persons perform an unpleasant behavior, their feeling of choice about whether or not to perform it may be varied. After the person has committed himself to performing the unpleasant behavior, he may be told that due to a chance factor, he will be deprived relative to other persons who perform the same behavior. The amount of relative deprivation can be varied. We might then expect that under low choice conditions, increasing relative deprivation would result in decreasing satisfaction. Our major hypothesis thus is that under greater but not under lesser degrees of choice, increasing relative deprivation tends to result in increasing satisfaction. Thus if chance relative deprivation affects dissonance only under conditions of prior choice, there should be an interaction between amount of choice and

[1] This experiment was performed as part of the Yale Communications Research Program. It was supported by grants to Carl I. Hovland from the Bell Telephone Laboratories and from the Rockefeller Foundation.

[2] Now at Duke University.

amount of relative deprivation in determining the amount of subsequent satisfaction.

METHOD

Subjects (*Ss*) were told that they must take part in a disagreeable and time-consuming experiment. Approximately half of the *Ss* were given a high opportunity, and half a low opportunity, to get themselves excused; within these groupings, half were placed in a situation of high relative deprivation and half in one of low relative deprivation. Immediately thereafter a questionnaire was administered to measure the *Ss*' attitudes toward participating in the experiment and to check the success of the experimental manipulations.

Subjects

Five sections of the introductory psychology course at Yale University were used. The *N*s were: 24, 24, 28, 29, and 29. *Ss* were approached in their classrooms during a regularly scheduled class period.

Procedure

The experimenter (*E*) was introduced as a member of the psychology faculty who was on a departmental committee concerned with student participation in psychological research. *E* explained that a new project had just been set up in which all students in introductory psychology classes were to participate. He said that unfortunately, this particular project wasn't going to be very interesting. He described the task as three solid hours of copying random numbers in order to establish norms for other research. He emphasized that it would be boring, tedious, no challenge, since they must not try to be extra fast or accurate in their copying, and not a very good opportunity to learn about psychological research.

Manipulation of Choice. *Ss* were then given schedule sheets on which to indicate their free hours. In the *Low Choice* condition, these sheets were collected without mention of any way for the *Ss* to be excused. In the *High Choice* condition, *Ss* were told that though it was desirable that everyone participate, a person could be excused if necessary. *E* told them that to be excused, they must indicate on a separate form attached to their schedule sheets their desire for an appointment with the Psychology Department's Director of Undergraduate Studies concerning their excuse. Two classes were given the High Choice manipulation, three were given the Low Choice manipulation.

Manipulation of Relative Deprivation. When the schedule sheets had been collected, the *Ss* were asked to evaluate the general issue of participation in research. Attached to each questionnaire was a page of "evaluation sheet instructions." These instructions, intended to induce the conditions of relative deprivation, said:

> In this research project you are to copy random numbers in an "average" way. You must not try to be unusually accurate or fast. A large body of evidence shows that rewards don't make much difference in this kind of research, and they probably won't be too important here either. Nevertheless, looking at the effects of rewards might not be a bad idea. Thus, while most persons will receive $\$____$ for their three hours of copying, a few, randomly chosen people will receive nothing. The following check mark indicates whether or not you personally will receive payment.
>
> Will receive payment_____
> Will not receive payment__X__

In the *Low Relative Deprivation* condition, the *Ss* were told that "most persons will receive $1.00" for their copying. In the *High Relative Deprivation* condition the *Ss* were told that "most persons will receive $10.00" for their participation. Half of each class were assigned to the high condition and half to the low condition. In both conditions, *Ss* were told that they would receive nothing. It should be noted that each person was told he was relatively deprived by *chance*.

In manipulating deprivation, it was necessary to keep the differential reward plausible, though not too plausible. To the degree that there were good reasons for others getting money, dissonance would be reduced. On the other hand, an entirely implausible reason would militate against the seriousness of the entire enterprise.

After reading the above "information," the *Ss* completed the questionnaire. The central dependent variable of the present study concerns the degree of satisfaction experienced by the *Ss* at the prospect of participation in the disagreeable research project. The degree of satisfaction was measured by the *Ss*' ratings on a nine-point scale which asked them the following: "Considering what you will get out of the experiment (interest, enjoyment, learning, etc.) how satisfied are you with having been assigned to this research?" Checks on the experimental inductions were also made. For the choice variable, the *Ss* were asked on a nine-point scale, "How difficult do you think it would be for you to get out of this experiment?" As a check on the relative deprivation inductions, the *Ss* were asked whether or not they would receive anything for their participation and whether or not they felt that they should receive anything.

The questionnaires were then collected and the *Ss* were apprised of the nature of the experiment. They were assured that they did not have to take part in the "research," and through an illustrative talk on experimental design, whatever annoyance or anxiety they had was presumably dissipated.

RESULTS

Effectiveness of the Experimental Manipulations

The data indicate that the manipulations of relative deprivation were entirely effective in creating perceptions of the differing amounts of money to be earned by others in the same class.

The variable of choice, however, presents another picture. In examining the *Ss*' perceptions of the degree of choice they had in submitting to the *fait accompli*, it was found that more variance in choice existed within experimental conditions than between them. In

TABLE 1
MEAN RATINGS OF SATISFACTION FOR EACH EXPERIMENTAL SUBGROUP

	Perceived Degree of Choice	
	Low	High
Replication #1 (N = 29)		
Low Rel. Dep.	5.19[a] [9]	4.00 [6]
High Rel. Dep.	3.82 [8]	5.50 [6]
Replication #2 (N = 24)		
Low Rel. Dep.	2.57 [7]	2.12 [5]
High Rel. Dep.	2.06 [8]	3.10 [4]
Replication #3 (N = 28)		
Low Rel. Dep.	4.81 [7]	3.04 [7]
High Rel. Dep.	2.58 [8]	3.78 [6]
Replication #4 (N = 24)		
Low Rel. Dep.	3.77 [3]	2.11 [10]
High Rel. Dep.	1.58 [6]	2.85 [5]
Replication #5 (N = 29)		
Low Rel. Dep.	4.94 [7]	3.07 [8]
High Rel. Dep.	3.84 [9]	2.34 [5]

[a] Small numbers indicate low satisfaction, large numbers, high.

effect, then, the manipulation of choice failed to determine the amount of choice felt by the *S*s. Our analyses therefore employ the degree of *perceived* choice to distinguish between those who experienced a high degree of choice and those who felt that there was little choice in submitting to the disagreeable *fait accompli*. Since the variances in choice differed from class to class, each class was dichotomized on choice as near as possible to the median, and each such subgroup further divided into the manipulated conditions of High and Low Relative deprivation. Thus we have five independent replications of an experiment where perceived choice varies and where relative deprivation varies but where all *S*s expect to participate in a disagreeable research project.[3]

Effects of Choice and Relative Deprivation on Satisfaction

Ratings of the *S*s' attitude toward the research assignment constitute the data of the present experiment. In Table 1 are presented the mean satisfaction ratings for each of the four conditions for all five replications. Corresponding *N*s are presented in parentheses.

[3] There are no data on expectation for participation in the conditions where Low Choice was manipulated; we must therefore assume that all *S*s perceived themselves to be committed.

TABLE 2
INTERACTION DIFFERENCES BETWEEN HIGH AND LOW RELATIVE DEPRIVATION CONDITIONS FOR EACH CHOICE CONDITION

Replication	Low Choice	Interaction: High Choice	Difference	p Value*
1	1.37	−1.50	2.87	.04
2	.51	−.98	1.49	ns
3	2.23	−.74	2.97	.07
4	2.19	−.74	2.93	.06
5	1.10	.73	.37	ns

* All p values are from two-tailed t tests.

The results are highly consistent. Except for a reversal of means in the High Choice condition in Replication 5, all the data support the assumption of an interaction between degree of choice and degree of deprivation in determining attitudes toward a prospective event. These results are summarized in Table 2 where the differences between deprivation conditions within the choice conditions and the interaction differences are presented. In the first two columns of Table 2, a positive figure indicates that those *S*s in the Low Relative Deprivation condition were more positive in their attitudes, a negative figure indicates that the High Relative Deprivation *S*s were more positive.

Table 2 shows that in all five replications the interaction differences are in the same direction. In one of the two cases where the difference is statistically unreliable (Replication 5) it is due to a reversal in the High Choice condition. In general, however, we may conclude that attitudes toward the unpleasant *fait accompli* are strikingly different depending upon degree of perceived choice and degree of relative deprivation experienced after commitment.

DISCUSSION

If we can assume that the present results would have been obtained had the manipulation of choice been successful, then our major hypothesis appears to have been supported. When an individual feels he has little prior choice in whether or not he is committed to an event, a direct effect of deprivation occurs: increasing relative deprivation results in decreasing satisfaction. If an individual feels he has at least some prior choice, increasing rela-

tive deprivation results in increased dissonance and consequently in increased satisfaction.

Since *S*s in the present experiment thought they were relatively deprived as the result of chance, certain kinds of events not under the person's control appear capable of producing dissonance. At the same time, a limit to these conditions is found. The data indicate that only when the person feels he has had some prior choice can a chance event affect the magnitude of his dissonance.

A second point of theoretical interest concerns the qualification of Merton's concept of relative deprivation. In the present data his analysis may apply only when the individual has had no choice concerning his assignment.

It should be noted that although relative deprivation was induced in the present experiment, any negative chance event introduced after commitment in a situation of High Choice would have the same theoretical status. Thus variations in the degree of any type of punishment might be expected to create similar variations in dissonance and consequent attitude change. In addition, as far as dissonance theory is concerned, differential relative deprivation ought to produce differential dissonance even in cases in which persons receive some rewards for engaging in the negative behavior. In the present experiment the *S*s varied in the degree to which they were punished relative to others. It would also seem that under conditions where forced compliance has been elicited to a disagreeable behavior, variations in the degree to which others are punished relative to the individual and variations in the degree to which he is rewarded relative to others should also produce variations in dissonance. In all instances, the discrepancy between his cognitions supporting his behavior and those negating it, would vary with degree of relative deprivation. The emphasis on relative deprivation highlights a view of motivation for compliance as not absolute but as relative to the situation of others around the individual. And it would follow that the more important or similar or salient those others, then the greater would be the effect on dissonance of variations in relative deprivation.

Unexplained Effects. While our original hypotheses account for some of the obtained effects, they do not account for all. We did not expect that persons in the condition of Low Relative Deprivation would tend to be less satisfied as their opportunity to be excused increased. There are several possible explanations of this apparent effect, of which we will mention one that seems plausible to us.

This explanation assumes that as the magnitude of force to perform an unpleasant behavior increases, a person tends less and less to try to reduce his dissonance by attempts to avoid its performance. In other words, the greater the force to comply, the more dissonance tends to be reduced by changes in attitude rather than by contemplation of avoiding compliance. Persons in a situation of Low Choice, compared to those in one of High Choice, should be more likely to reduce their dissonance by increases in satisfaction. In general, therefore, those confronted with High Choice might tend to show less satisfaction. This result would not, of course, alter the fact that more dissonance may be created in the condition of High Choice than in that of Low Choice, so that the obtained interaction may still be expected to occur.

Implications for Further Research. While the present data suggest that a chance event can affect the magnitude of dissonance if a person has made a prior relevant choice, this interpretation is somewhat equivocal because of the failure of the choice manipulation. A true experimental demonstration that a chance event can affect dissonance is still needed.

A second ambiguity of interpretation centers on the fact that persons in the condition of Low Relative Deprivation tended to be less satisfied as their opportunity to be excused increased. This paradoxical phenomenon deserves thorough investigation. One possible line of research might follow the explanation suggested above: that increasing the force to comply decreases one's ability to reduce dissonance by contemplation of escape from compliance. One might expect, for example, that this would occur where coercion is used, as in the present example, but would not occur where the inducing force was a reward. Other possible explanations must also be systematically investigated experimentally.

SUMMARY

This experiment was designed to explore some of the limiting conditions under which cognitive dissonance and consequent attitude change are produced. Two questions were asked: (*a*) can a chance event affect the magnitude of dissonance and (*b*) does the effect of such a chance event depend upon there having been a prior choice in commitment to an event?

In order to explore these hypotheses, perceived choice in submitting to an unpleasant event and a chance relative deprivation after commitment was varied. Five independent replications were carried out in five introductory psychology sections. The *S*s were asked to and did commit themselves to participation in a disagreeable and time-consuming project. They varied in their perception of how much choice they had in getting out of it if they wanted to, and in the degree to which relative deprivation in connection with their participation was introduced after commitment.

The results indicate an interaction such that under conditions of low perceived choice, the more relative deprivation *S*s suffer, the more *negative* are their attitudes toward the *fait accompli*, while under conditions of high choice, the more relative deprivation they suffer, the more *positive* are their attitudes toward the *fait accompli*. Thus, a chance event may be seen to affect the magnitude of dissonance and consequent attitude change but only under conditions of high choice.

The results were discussed in terms of the bearing of dissonance-producing chance events and relative reward and punishment on dissonance theory. Alternative explanations for the results were also offered.

REFERENCES

BREHM, J. W. Behavioral compliance and attitude change. Depart. of Psychol., Yale Univer., 1957. (Mimeo.)

BREHM, J. W. Increasing cognitive dissonance by a *fait-accompli*. *J. abnorm. soc. Psychol.*, 1959, **58**, 379–381.

COHEN, A. R., BREHM, J. W., & FLEMING, W. H. Attitude change and justification for compliance. *J. abnorm. soc. Psychol.*, 1958, **56**, 276–278.

FESTINGER, L. *A theory of cognitive dissonance*. Evanston: Row Peterson, 1957.

FESTINGER, L., & CARLSMITH, J. H. Cognitive consequences of forced compliance. *J. abnorm. soc. Psychol.*, 1959, **58**, 203–210.

MERTON, R. K., & KITT, ALICE. Contributions to the theory of reference group behavior. In R. K. Merton & P. F. Lazarsfeld (Eds.), *Studies in the American soldier*. Glencoe: Free Press, 1950.

Received March 19, 1958.

ATTITUDINAL EFFECTS OF CHOICE IN EXPOSURE TO COUNTERPROPAGANDA[1]

ARTHUR R. COHEN, HERBERT I. TERRY, AND CHARLES B. JONES
Yale University

IN A series of experiments stimulated by Festinger's (1957) theory of cognitive dissonance it has been found that when a person is induced to take a behavioral stand contrary to his cognitions he may reduce the dissonance or tension between his behavior and cognitions by making those cognitions (i.e., opinions) consistent with the behavior (Brehm: 1957a, 1957b; Cohen, Brehm, & Fleming, 1958; Festinger & Carlsmith, 1959). The theory also states that the introduction of information or opinions discrepant from the individual's own opinion also produces dissonance and consequent pressure to reduce that dissonance. However, it is not entirely clear from the theory whether simple exposure to discrepant information is enough to produce dissonance and consequent attitude change in line with the information, or whether or not a more definite behavioral stand on the part of the individual to expose himself to such information is necessary. Research (Brehm, 1959; Brehm & Cohen, 1959) has shown that chance events may produce cognitive dissonance and consequent attitude change without the person's taking a stand relative to that particular event. It appears, though, that at least *some prior decision* in which the person commits himself to the general situation is necessary. The present experiment is designed to study further the effects of choice on cognitive dissonance.

The theory states that under certain conditions exposure to contrary information produces dissonance between the person's prior cognitions and his new cognitions. Everything else being equal, we might expect that the greater the difference between the person's opinion and the new information, the greater the dissonance produced. The dissonance produced by exposure to new information may be reduced by misperception, misinterpretation, rejecting those who disagree, changing one's own opinion, influencing others, and so forth (Festinger, 1957). However, dissonance reduction through change in opinion to fit the new information should be most likely to result when the person has in some active way chosen to expose himself to the information. Thus, under conditions where he has had little choice in exposure, a direct effect of persuasion might be expected to occur: the greater the discrepancy between his prior opinion and the new information, the greater his resistance to changing his attitudes in line with the new information. Where the individual has chosen to expose himself to counterpropaganda, greatest dissonance should be expected where his own opinions are most different from the new information (Brehm, 1959). The discrepancy between his own opinion and the behavioral stand involved in his decision to entertain counterpropaganda can be resolved by changing his opinions toward the new information. Our major hypothesis, thus, is: *Under greater but not under lesser degrees of choice, increasing the discrepancy between a person's initial opinion and new information confronting him gives rise to increasing dissonance and consequent attitude change.* Thus, if the difference between the individual's initial opinion and the new information affects dissonance only under conditions of prior choice, there should be an interaction between amount of choice and amount of discrepancy in determining the degree to which attitude change occurs.

METHOD

Subjects (*Ss*) were asked to take part in a "survey." They were first given a prequestionnaire on the attitude dimension, then given the major experimental manipulations. Half were given a high opportunity to avoid hearing counter-propaganda, and half a low opportunity. They were then exposed to the new information. Immediately thereafter a questionnaire was administered to measure the *Ss*' attitudes on the same dimension and to check the success of the experimental inductions.

Subjects

Seventy-one freshmen at Yale University comprised the sample. They were approached individually in their dormitories.

[1] This report was prepared as part of the Yale Communications Research Program.

Procedure

The experimenter introduced himself to each *S* as someone taking a survey for the "Department of Human Relations." The survey concerned the advisability of young men marrying before the age of 23 (about two years after graduation). The issue was a then current topic of discussion in the campus newspaper and among the undergraduates, and was thus assumed to be salient.

Ss first filled out an a priori six-point scale as to how they felt about "men marrying by the age of 23." The scale ran from "agree very strongly" to "disagree very strongly." When the *S* had completed the questionnaire, *E* noted which side of the attitude continuum the *S* was on and chose the communication that opposed the *S*'s opinion. In all cases *Ss* received new information *contrary* to their initial opinions.

Manipulation of Choice. When the opinion measures had been collected *E* said to all *Ss*, "I have here some additional information on early marriage collected by the Department of Human Relations from the results of previous surveys, which I would like to read to you now. It does, however, uphold a point of view quite contrary to the one you have taken." In the *High Choice condition*, the *Ss* were told: "I'll leave it entirely up to you whether or not I read it to you. Would you like to hear it?" If *S* said yes, he was asked, "Are you sure?" If he said yes again, he was asked again if he was sure. If he still agreed, *E* then read the appropriate argument to him. In the *Low Choice condition*, *Ss* were merely told that *E* was going to read the contrary information to them.[2]

The Information. The issue of early marriage was chosen because of its salience and because it was relatively easy to construct arguments for each side of the issue. Both arguments contained some facts and some half-truths. In order to equate the strength of the opposing communications, arguments used on one side were merely negated and used on the other side. Evidence on the success with which arguments were equated will be presented later.

Discrepancy Between Initial Opinion and New Information. This variable was not manipulated experimentally but was controlled through the initial measures of the *Ss*' opinions. *Ss* who were extreme on the attitude scale on either one side of the scale or the other were assumed to represent the condition of *High Discrepancy*; *Ss* whose responses fell in the middle of the attitude scale were assumed to represent the condition of *Low Discrepancy*. With arguments on a given side constant, the more extreme the individual is in his attitude, the greater the discrepancy between his opinion and the new contrary information. In addition to noting which side of the issue *S* was on before introducing the information, *E* also noted the extremity of his opinion in order to assure equal and random assignment into the discrepancy conditions.

Postmeasures. After hearing the new information, *Ss* filled out a postquestionnaire. They were told, "Now that you have heard this additional information on early marriage, compiled from the results of previous surveys, we would like to get some of your reactions to it. You may possibly want to evaluate the situation in the light of this information. So would you please fill out this questionnaire." *Ss* filled out an opinion scale identical with the premeasure and also an item checking on the choice manipulation. On the latter they were asked, "Considering that you agreed to participate in this research, how much choice do you feel you had in whether or not you were given the additional information on early marriages?" The scale used was a six-point scale which ranged from "complete choice" to "no choice at all."

Ss were then told in some detail about the nature and true purpose of the experiment and cautioned not to speak to others about it for an appropriate time.

RESULTS

Effectiveness of the Experimental Manipulations

The data indicate that the choice manipulations were entirely effective in creating perceptions of differing degrees of choice in exposure to the new information. The choice conditions show extreme differences in the expected direction: the mean for the High Choice condition is 1.96; for the Low Choice condition, 4.36. This difference is significant beyond the .001 level ($t = 9.21$).[3]

The data also show that the information given to the *Ss* was fairly convincing and that the opposing arguments were of approximately equal strength. When the change in opinion from before to after measurement is taken for the entire population, the average change is .69, which is significantly different from no change ($p < .05$). Furthermore, the change in opinion among those *Ss* who were pro and received con arguments is no different from the change among those who were con and received pro arguments. The average change among the former is .74, among the latter .68. Thus, we may safely assume that differences between our experimental subgroups need not be due to

[2] Since all *Ss* had already filled out a prequestionnaire, practically all *Ss* assigned to the High Choice condition agreed to hear the opposing arguments: only three *Ss* (with High opinion discrepancy) refused to entertain the counterpropaganda. These *Ss* were eliminated from the experiment. However, the deletion of these *Ss* is not enough to seriously bias the results. Furthermore, the fact that three *Ss* felt enough choice to refuse indicates that constraints against not listening to the opposing arguments were not overly great and that there was in fact high choice in this condition. Further evidence on the effectiveness of the choice manipulation will be given with the results.

[3] All statistical tests reported here are two-tailed tests.

TABLE 1
MEAN OPINION CHANGE FOR EACH EXPERIMENTAL SUBGROUP

Discrepancy Between Opinion and Information	Degree of Choice	
	Low	High
Low	.76 [19]	.63 [20]
High	.32 [17]	1.17 [15]

differences in the strength of the communications.

Effects of Choice and Discrepancy on Attitude Change

The change in *S*s' opinions toward the marriage issue from before to after exposure to new counterinformation constitute the data of the present experiment. In Table 1 are presented the opinion change scores for each of the experimental subgroups. Corresponding *N*s are in parentheses; the higher the figure, the more the change.

In Table 1 it can be seen that the data support the assumption of an interaction between degree of choice in exposure and degree of opinion discrepancy in determining attitude change toward new discrepant information. The interaction t equals 2.49, $p < .02$.

DISCUSSION

In the present experiment, the discrepancy between initial opinion and new information was not manipulated experimentally but was allowed to vary with the *S*s' own opinion preferences. This procedure could allow some ambiguity of interpretation if self-selection into the choice conditions had been permitted or if initial opinions were highly correlated with attitude change. But since degree of choice was experimentally manipulated and there are no reliable over-all differences in opinion change between the High and Low Discrepancy conditions, we can assume that our major hypothesis has been supported. Thus, when a person feels that he has little prior choice in exposure to information counter to his opinion, increasing the discrepancy between his opinion and the information results in decreasing attitude change. If he feels he had at least some prior choice, increasing the discrepancy results in increased dissonance and consequent attitude change.

Under conditions of Low Choice, the relationship between discrepancy and attitude change may be conceived of as one of direct persuasion. The more extreme a person is in his own opinion the more he may be expected to resist counterarguments. Because he has had no choice in exposure, there is little dissonance between his cognitions (initial opinion) and behavior. What dissonance there remains between his opinion and the opposing information he can reduce in ways other than attitude change. However, when a person has chosen to expose himself to the countercommunication, dissonance is produced between his cognitions (initial opinion) and behavior (choice of listening to discrepant information). In addition, the more discrepant the information, the more dissonance should be produced. In a situation in which he has chosen to take a stand contrary to his cognitions, the dissonance can be reduced by making those cognitions consistent with the behavior (i.e., change his attitude to conform with the new information). And we would expect, the more the dissonance, the greater the pressure to reduce it and, therefore, the greater the consequent attitude change.

Though the original hypothesis accounts for most of the obtained effects, it may not account for all. It was not expected that in the Low Discrepancy condition, choice would make little difference in attitude change. An alternative explanation would assume that where no potentiality for dissonance exists as a result of the small discrepancy between initial opinion and new information, choice in exposure is not seen as a stand sharply etched against one's own cognitions. As the discrepancy increases, the decision to listen to counterinformation is seen as a behavioral commitment quite contrary to one's own position, resulting in increased dissonance and consequent attitude change. A more thorough exploration of this issue through the experimental manipulation of opinion discrepancy is still necessary.

Nevertheless, it may be said that exposure to counterinformation per se may not be enough to produce dissonance, but may only mobilize the usual methods of dealing with attempts at persuasion, depending upon how different from his own position the person sees the new material to be. In situations in which

he has actively chosen to expose himself to discrepant information, however, dissonance may result and lead to reduction through attitude change in line with the new information.

SUMMARY

This experiment explored some of the conditions under which exposure to information discrepant from one's own opinion produces cognitive dissonance and consequent attitude change. It was assumed that chance exposure to such information would produce little dissonance, but that the person's choice to expose himself to such information would produce high dissonance and consequent attitude change, the greater the discrepancy the greater the dissonance.

To test this hypothesis, choice in exposure to contrary information was varied, and the person's opinion on a salient attitude dimension was measured. *Ss* were approached individually and a premeasure of their attitude was taken. Half the *Ss* who were extreme in their opinions and half the *Ss* who were moderate were simply exposed to information counter to their opinions; the other half of these groups were given a choice in exposing themselves to such information. After exposure, their opinions were measured again.

The results indicate an interaction such that under conditions of *Low Choice*, a direct persuasion effect occurred: *Ss*' attitudes toward the issue changed more in line with the new information, the *less* the discrepancy between the information and their initial opinion. However, under conditions of *High Choice*, expectations from dissonance theory were fulfilled: *Ss*' attitudes changed more, the *more* the discrepancy between the information and their initial opinion. The results were discussed in terms of the importance of prior choice in exposure in creating dissonance.

REFERENCES

BREHM, J. W. Forced verbal compliance and attitude change. Depart. of Psychol., Yale Univer., 1957. (Mimeo.) (a)

BREHM, J. W. Behavioral compliance and attitude change. Depart. of Psychol., Yale Univer., 1957. (Mimeo.) (b)

BREHM, J. W. Increasing cognitive dissonance by a *fait accompli*. *J. abnorm. soc. Psychol.*, 1959, **58**, 379–382.

BREHM, J. W., & COHEN, A. R. Choice and chance relative deprivation as determinants of cognitive dissonance. *J. abnorm. soc. Psychol.*, 1959, **58**, 383–387.

COHEN, A. R., BREHM, J. W., & FLEMING, W. H. Attitude change and justification for compliance. *J. abnorm. and soc. Psychol.*, 1958, **56**, 276–278.

FESTINGER, L. *A theory of cognitive dissonance*. Evanston: Row Peterson, 1957.

FESTINGER, L., & CARLSMITH, J. H. Cognitive consequences of forced compliance. *J. abnorm. and soc. Psychol.*, 1958, **58**, 203–210.

Received March 20, 1958.

EVALUATION OF THERAPY BY SELF-REPORT: A PARADOX[1]

JANE LOEVINGER AND ABEL OSSORIO
Jewish Hospital, St. Louis, and St. Louis State Hospital

ROSE (1956) has recorded some data which suggest that soldiers whose responses on personality questionnaires are "normal" and indicate "high morale" are a little more likely to break down in battle than those who give more aberrant responses. Here is a paradox to which we should not blind ourselves by clinging to an arbitrary definition of normality. The idea of "normality" or "good adjustment" is a complex one that may contain mutually contradictory implications. Is the purpose of therapy always to increase "normality" or "good adjustment?"

Measurement and description of personality have at various times been stranded on two opposite shores of the main stream. On the one side are those who get lost in the particularities of the process they set out to describe: Here we find Erikson's (1950) seven stages in the normal process of ego development and Rogers' (1958) seven-stage description of the process of psychotherapy. However illuminating as descriptions, these intricate schemes do not lend themselves to psychometric assay.

On the other bank are the testers with their abstractions and oversimplifications: Everyone must be more or less of one thing, like "adjusted," or somewhere between a pair of poles, like "authoritarian-democratic." If one superposes Erikson's portrait of normal development, Rogers' portrait of psychotherapy, and the results of objective personality measurement, especially those pertaining to authoritarianism and response stereotypy, what emerges as from a composite photograph is a variable whose core is *ability to conceptualize oneself* as a psychological person.

Let us look first at this trait as an aspect of normal development. During the second or third year, the child first learns that he has impulses and that acting on them is in some cases forbidden and in other cases must be deferred. That this time is the point of origin of the moral sense has often been emphasized. Here we emphasize instead that it is the point of origin of one's conception of his inner life. To conceive of one's impulses is to achieve some measure of control over them. One of the major tasks of childhood is extension and elaboration of the concepts of impulse and control.

By early adolescence there is normally established a strict moral sense, at least on the verbal level. The typical adolescent is rigid, prone to think in stereotypes, intolerant of deviations, punitive, and antipsychological—in short, what has been called an authoritarian personality (Erikson, 1950; Remmers & Radler, 1957). His self-conception is conventional and expressed in "socially desirable" terms.

During the college years, particularly in sophisticated Eastern schools, there is a decline in authoritarian tendencies and a shift to a liberal and tolerant view of others and a more sophisticated, differentiated, self-critical, and realistic self-concept (Christie, Havel, & Seidenberg, 1958; Cohn, 1957; Sanford, 1956). Where the adolescent's standards are typically expressed, "The right way is . . .," the more mature person's standards are more likely to be expressed, "I am the kind of person who. . . ." In terms of the growth of the capacity for self-conception, we express it this way: The infant has no distance from his impulses; the adolescent has distance from impulse but not from ego; the mature person has capacity for distance from ego as well as impulse.

If we project these three stages onto an adult screen, we see that they correspond to meaningful individual differences. There are older children and even adults who have a degree of impulse control appropriate only to young children. Redl and Wineman (1957) have portrayed preadolescent children of this type. As adults they turn up as delinquents or psychopaths or disorganized personalities or as parents of delinquent children. They

[1] Preparation of this paper was supported in part by a research grant, M-1213, from The National Institute of Mental Health, Public Health Service.

come to social agencies and guidance clinics as involuntary referrals. However miserable their troubles, they "don't know why they were sent," and have nothing to say. They are unable to verbalize about themselves or even about their way of life. They come to an alcoholic clinic asking, "Where do I get the shots?" Their only conception of trouble is organic. It may take many weeks to bring such adults to the point of understanding what psychotherapy is about.

The differences between conventional-authoritarian adults and liberal-sophisticated adults are those which have been delineated by over a decade of research with the California F scale and similar tests. The great mass of the population is not much further along the scale of ego development than the typical adolescent. At the conclusion of the California study of the authoritarian personality, the authors came to realize that the psychopathic authoritarian and the conventional authoritarian are distinct types (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950, p. 972). They aimed to study the pathological type but ended by studying chiefly the conventional type. The pathological authoritarian appears to be drawn from what we have called the lowest level of ego development, the conventional authoritarian from our middle level. However, connotations of the term authoritarianism are too harsh, too pathological, and too gratuitously political for the present context; conventionality or even Philistinism is closer to characterizing our middle level.

If we define normality in statistical terms, the conventional person with a stereotyped self-concept is normal, whereas the more mature person with a more realistic self-concept is less normal because less numerous. On the other hand, if we take success in life as the criterion of good adjustment, clearly those who remain at the lowest level of ego development beyond infancy will be poorly adjusted in most instances. But there is probably little difference in life success between the conventional and the intellectually sophisticated. Elevation of the Philistine to places of power is a well-known hazard of democracy.

The conventional personality typically deals with unacceptable impulses by repression and denial. This is part of what the California group called "anti-intraception." Attempts to disseminate publicly the findings of psychology, particularly when they undermine the average person's mental habits, are likely to strengthen his antipsychological defenses. He is by no means exempt from psychological illness, but he is debarred from utilizing the services of psychotherapists by his antipsychological defenses. If he comes into psychotherapy, it will often be because of medical referral for psychosomatic illness or because of an acutely disabling condition (Hollingshead & Redlich, 1958, Ch. 6).

To come voluntarily for psychotherapy for a mildly disabling condition already implies some sophistication in one's conceptualization of his inner life, a capacity for detachment and self-criticism. In addition, all forms of psychotherapy operate by means of increasing one's capacity for realistic self-conception. At this point we are tracing the connection between our concept of ego development and Rogers' concept of the process of therapy. While we emphasize the cognitive aspect, and therefore speak of increased capacity for distance from ego in the highest stage, Rogers speaks of decreased distance from feelings in the final stages of therapy. There is no contradiction, however. It is partly a matter of how one turns the metaphor: Rogers means by distance *alienation*, whereas we mean *perspective*. The feelings which Rogers and other therapists teach their patients to perceive as feelings were not previously absent but were perceived as properties of the world rather than as feelings of the person. In *The Authoritarian Personality*, this characteristic was labelled "projectivity."

That therapeutic progress can be viewed both as greater distance from ego and closeness to feelings can be accounted for in terms of the psychoanalytic concept of the temporary splitting of the ego which occurs during therapy (Ekstein, 1956). In requiring verbalization of feelings, psychotherapy both requires and encourages some capacity for realistic self-conception.

Sullivan (1953) traces the origin of perspective on oneself to the interpersonal situation, i.e., communication with another person is the basis for learning to see oneself as others

see one. Therapy is an intimate situation between the patient and one other person, with the patient's inner life as subject for discussion. Surely this is an ideal setting in which to acquire the ability to see oneself as another person would.

The average person just does not have this ability, and it is fallacious to use his self-conception, as expressed in interview or personality test, to evaluate the success of therapy. Similarly, a high correlation between one's self-description and one's description of the ideal person is probably a better indication of Philistine self-satisfaction than of the success of therapy. The findings of several investigators that self-reports are inconsistent with or poorly correlated with other measures of adjustment (Beilin, 1957) and of success in therapy (Kelman & Parloff, 1957) are, we believe, clarified by these considerations.

In view of current interest in individuals who withdraw from therapy at the initial stages or who withdraw their children from child guidance clinics before treatment, we would like to record an hypothesis concerning some proportion of those cases. The postulated sequence of events is that an acutely stressful occurrence leads to temporary breaching of the conventional individual's antipsychological defenses, leading him to seek help in psychotherapy. The antipsychological defenses heal over before it is possible to gain access to the underlying difficulty. In such cases, one would expect on follow-up to get bland, "normal," socially approved self-reports on test and interview, regardless of the outcome of the disturbances which led them to register for treatment. By contrast, the self-critical reports of those who stay in therapy are not a fair indication that therapy has not succeeded.

Scott (1958) has ably summarized the definitions of mental health and mental illness utilized in current research. Although the theme of his paper is different from that of the present note, certain conclusions could be justified from his considerations as well as from our own.

It is by no means obvious where one can find either a normal control group to epitomize the desired outcome of therapy or a neurotic control group to serve as baseline for evaluating therapeutic success. Verbal self-reports, particularly those elicited by objective personality tests, and possibly also those elicited by brief interviews, are not satisfactory data for evaluation of therapy. They reflect too much the conventionality-sophistication dimension, too little other more important aspects of adjustment.

We have no easy solutions to offer. But it seems better to acknowledge the difficulties of the problem than to draw misleading conclusions from spuriously objective criteria of therapy.

REFERENCES

ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.

BEILIN, H. The prediction of adjustment over a four year interval. *J. clin. Psychol.*, 1957, **13**, 270–274.

CHRISTIE, R., HAVEL, JOAN, & SEIDENBERG, B. Is the F scale irreversible? *J. abnorm. soc. Psychol.*, 1958, **56**, 143–159.

COHN, T. S. The relation of the F-scale to intelligence. *J. soc. Psychol.*, 1957, **46**, 207–217.

EKSTEIN, R. Psychoanalytic techniques. In D. Brower & L. E. Abt (Eds.), *Progress in clinical psychology*. New York: Grune & Stratton, 1956. Pp. 79–97.

ERIKSON, E. H. Growth and crises of the healthy personality. In M. J. E. Senn (Ed.), *Symposium on the healthy personality*. New York: Macy Found., 1950.

HOLLINGSHEAD, A. B., & REDLICH, F. C. *Social class and mental illness*. New York: Wiley, 1958.

KELMAN, H. C., & PARLOFF, M. B. Interrelations among three criteria of improvement in group therapy: Comfort, effectiveness, and self-awareness. *J. abnorm. soc. Psychol.*, 1957, **54**, 281–288.

REDL, F., & WINEMAN, D. *The aggressive child*. Glencoe, Ill.: Free Press, 1957.

REMMERS, H. H., & RADLER, D. H. *The American teenager*. Indianapolis, Ind.: Bobbs-Merrill, 1957.

ROGERS, C. R. A process conception of psychotherapy. *Amer. Psychologist*, 1958, **13**, 142–149.

ROSE, A. M. Conscious reactions associated with neuropsychiatric breakdown in combat. *Psychiatry*, 1956, **19**, 87–94.

SANFORD, N. (Ed.) Personality development during the college years. *J. soc. Issues*. 1956, **12**, 1–71.

SCOTT, W. A. Research definitions of mental health and mental illness. *Psychol. Bull.*, 1958, **55**, 29–45.

SULLIVAN, H. S. *The interpersonal theory of psychiatry*. New York: Norton, 1953.

Received October 10, 1958. Prior publication.

CRITIQUE AND NOTES

DISPLACEMENT OF PREFERENCE AS A FUNCTION OF AVOIDANCE-AVOIDANCE CONFLICT[1]

HAVA BONNÉ GEWIRTZ

University of Maryland

AN EARLIER study (Gewirtz, 1959) has demonstrated that approach and avoidance gradients of preference could be obtained from children, following positive or negative reinforcement (success or failure) applied to their response to a training task. The present experiment represents an extension of that study, and attempts to establish a condition under which preference *displacement* may occur.

Displacement, as the term is used here, describes the phenomenon in which responses shift from the conditioned stimulus towards less similar stimuli, as a function of conflicting tendencies associated with the conditioned stimulus. Miller (1944, 1948) has used the concepts of stimulus generalization (approach and avoidance gradients) in developing his theory of conflict and displacement. While "approach-avoidance" conflict is his major model for displacement, he has also advanced the possibility of displacement resulting from "avoidance-avoidance" conflict. In this case, responses to stimuli at both extremes of a stimulus dimension have been negatively reinforced, producing two avoidance gradients opposite in direction. And when *S* is forced to respond (i.e., when there is no alternative to responding), it is expected that his response would be displaced to a stimulus corresponding to the position where the two avoidance gradients intersect. If these gradients are assumed to be of equal strength at their respective points of origin, the point of their intersection would correspond to the middle of the stimulus dimension. Thus, the resulting displacement curve is expected to have its peak around the central stimulus on the dimension employed.

The present study represents an adaptation of this avoidance-avoidance conflict model to a children's play situation. Its first phase, referred to as the *Avoidance* experiment, is designed to ascertain the shape of the response pattern generated by the negative reinforcement condition employed. In its second phase, referred to as the *Displacement* experiment, the conflict treatment is applied; and it is expected that response displacement will result.

[1] This study was done while the author was at the University of Chicago.

METHOD

Since the method employed in this study is generally identical with that of the earlier generalization study (Gewirtz, 1959), it is described here only in brief. Except when otherwise stated, the conditions described below apply equally to both Avoidance and Displacement experiments.

Subjects

Twenty children (8 boys and 12 girls) between the ages of six and eight years served as *S*s. They were selected at random from the first and second grades of a private school in the Chicago area. In terms of socio-economic background, achievement, and intelligence level, this sample was considered similar enough to the one employed in the generalization study to permit relevant comparisons between some of the experimental results obtained from the two studies.

Materials

Five formboard-type puzzles were constructed in such a way that their shapes constituted a similarity dimension, ranging from a square diamond-shaped groove to a long and narrow diamond-shaped one. The two puzzles representing the extremes of the dimension, i.e., Puzzle 1 and Puzzle 5, were used as training puzzles. Each was equipped with a set of seven pieces which could be fitted into the groove. But while this task appeared simple, the asymmetry of the pieces made solution extremely difficult—and actually impossible—for children of the age group employed.

Experimental Treatment

The experimental situation involved a puzzle-solution game played in the presence of the experimenter (*E*). The treatment condition consisted of negative reinforcement applied to *S*'s attempts to solve the assigned training puzzle, when solution was given up by *S* either spontaneously or following *E*'s suggestion (this was considered as one trial). Negative reinforcement represented a combination of (*a*) actual failure in puzzle solution, (*b*) mild verbal reproof by *E*, and (*c*) withdrawal of a small prize (plastic trinket) which had been promised earlier as a reward for successful solution.

General Procedure

*S*s were taken individually from their classrooms for the period of the experiment, which lasted approximately 20 minutes. In order to acquaint *S* with the similarity dimension, he was first shown the five puzzles, arranged in a row in random order, and was asked to

TABLE 1
PREFERENCE PROPORTIONS FOR AVOIDANCE AND DISPLACEMENT GROUPS

Group	N	Preference Proportions: Puzzles 1	2	3	4	5	Tests of Equality[a]: χ^2_{4df}	p
Avoidance	10	.02	.04	.06	.20	.68	52.29	< .001
Displacement								
Sequence 1–5	5	.13	.30	.47	.09	.02	—	.0003
Sequence 5–1	5	.10	.32	.40	.13	.05	—	.0083
Combined	10	.12	.31	.44	.10	.03	36.61	< .001

[a] When $N = 5$, p values were obtained directly from available tables (Bradley & Terry, 1952).

rearrange them in terms of their similarity to each other. The experimental treatment followed this introduction, and is described below separately for the Avoidance and for the Displacement experiments. Following the experimental treatment, *E* presented all five puzzles to *S* by the paired comparison method, and asked him, after each of the 10 pair presentations, to indicate which of the two puzzles in each pair he preferred to play with most. The differential preference rankings assigned to the five puzzles constituted the response measure.[2]

Avoidance experiment procedure. Ten *S*s were used in the experiment. Each was given a single negative reinforcement trial on Puzzle 1.

Displacement experiment procedure. Ten *S*s were used in this experiment. Each was given two negative reinforcement trials, one with each training puzzle (1 *and* 5). After the first trial, *S* was urged to try and solve the other, "different," puzzle. In order to control for possible sequence effects, five *S*s had their first trial with Puzzle 1 and five *S*s had their first trial with Puzzle 5.

RESULTS

The composite response of each treatment group was analyzed by means of the rank analy. method (Bradley & Terry, 1952), which enable assessment of the differences among the preference rankings assigned to the five puzzles (tests of puzzle preference equality), as well as assessment of differences in response pattern between groups of *S*s (tests of agreement). Table 1 presents the composite preference proportions (derived from the paired comparison rankings) for the five puzzles, both under the Avoidance and under the Displacement conditions, and the results of the respective tests of puzzle preference equality.

Avoidance Experiment

As seen in the first row of Table 1, the negative reinforcement applied to solution attempts with Puzzle 1 generated a preference pattern which deviated significantly from one where all puzzles are preferred equally. Preference for the puzzles increased with their increasing dissimilarity to the training puzzle, indicating that a steep and consistent avoidance gradient was produced.

For the purpose of the present study, it is also necessary to postulate that a similar avoidance gradient, opposite in direction, would result from a negative reinforcement trial applied to the opposite extreme of the dimension, i.e., to Puzzle 5. Since this assumption appears strongly supported by the findings of the Generalization study,[3] it was considered reasonable to assume that the avoidance gradient found here, if reversed in direction, would represent adequately the hypothetical gradient which would be generated by the same treatment when applied to Puzzle 5.

Displacement Experiment

Group data. The data were analyzed first separately for each of the two sequence subgroups (with five *S*s in each). As shown in Table 1, each subgroup responded to the treatment condition with differential preferences for the five puzzles which deviated significantly from equality. A test of agreement indicated no significant differences ($\chi^2 = 1.36$, $p > .80$, 4 *df*) between the two subgroup patterns, showing that order of trials with the two training puzzles did not affect the response. Hence, the subgroups' preference rankings were combined for all further analyses. The preference proportion values obtained from this Combined Displacement group (as well as sepa-

[2] According to the scoring method employed, the preferred member in the pair of puzzles received the rank of 1, and the rejected member received the rank of 2. Each of the five puzzles appeared a total of four times in the 10 pair presentations. Hence, for a single *S*, the sums of ranks for the five puzzles ranged between the values of 4 and 8, i.e., for the most and least preferred puzzle, respectively. Composite (group) scores were based on group averages of such individual sums of ranks.

[3] It was found that (*a*) negative reinforcement (three trials) applied to Puzzle 5 generated a reliable avoidance gradient, opposite in direction and similar in slope to the one generated by the identical treatment applied to Puzzle 1; and (*b*), that the three trials of negative reinforcement applied to Puzzle 1 produced a reliable avoidance gradient which was very similar to the one generated by a single trial of negative reinforcement on the same training puzzle. These comparisons suggest that the negative reinforcement treatment was more effective in determining the response pattern than was the particular training stimulus employed, or even the number of trials. Hence, the assumption above appears reasonable.

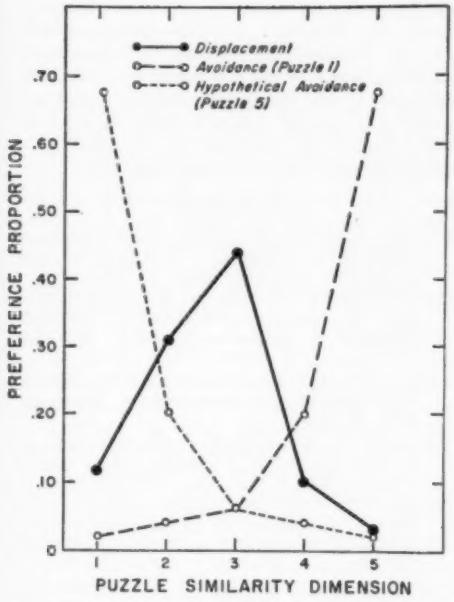


FIG. 1. COMBINED DISPLACEMENT CURVE, AVOIDANCE GRADIENT ORIGINATING AT PUZZLE 1, AND HYPOTHETICAL AVOIDANCE GRADIENT ORIGINATING AT PUZZLE 5 (N = 10 PER GROUP)

rately from each subgroup) indicate that Puzzle 3, the central puzzle on the dimension, was the most preferred, while Puzzles 5, 4, and 1 were the least preferred. These differential preferences constituted a displacement pattern which deviated significantly from equality. The Combined Displacement curve, the avoidance gradient generated by negative reinforcement applied to Puzzle 1, and the hypothetical avoidance gradient originating at Puzzle 5 (which is taken to be the mirror image of the Puzzle 1 avoidance gradient) are all illustrated in Fig. 1.

Individual data. More precise evidence for the occurrence of the displacement pattern was sought by an examination of *Ss*' individual response patterns. For this purpose, *S* was said to exhibit a "displacement pattern" when the values of his sums of ranks for Puzzles 1 and 5 were larger than those for the three intermediate puzzles (e.g., 8, 6, 4, 5, 7 or 7, 5, 5, 6, 7 for the five puzzles arranged in terms of the similarity dimension; see Footnote 2). A "perfect" individual displacement pattern was defined as one in which maximal preference was given to Puzzle 3 and minimal preference given to Puzzles 1 and 5 (i.e., sums of ranks representing the values 7, 6, 4, 6, 7). Table

TABLE 2

NUMBER OF *Ss* EXHIBITING DISPLACEMENT PATTERNS WITHIN THE DISPLACEMENT GROUP AND OTHER TREATMENT GROUPS

Group	Pattern		Difference from Displacement Group[a]	
	Displacement	Nondisplacement	χ^2_{1df}	p
Displacement	7	3	—	—
Avoidance	0	10	—	.002
Nonreinforcement[b]	5	15	—	.024
Positive[b]	5	35	11.52	< .001
Negative[b]	2	38	18.71	< .001

[a] When $N \leq 20$ in the group compared with the Displacement group, exact probabilities were obtained from available tables (Latcha, 1953).

[b] These three groups are from the generalization study (Gewirtz, 1959).

2 presents the distribution of displacement patterns within the Displacement group, the Avoidance group, and three other groups taken from the Generalization study.[4] One-tailed tests of independence were employed to compare the proportion of individual displacement patterns found within the Displacement group with that proportion found within each of the other four groups. As shown, the proportion of *Ss* exhibiting displacement patterns in the Displacement group (7/10) was significantly greater in each case than that proportion produced by a number of different, nonconflict treatments (0/10 in the Avoidance group; 5/20 in the Nonreinforcement group; 5/40 in the Positive group; and 2/40 in the Negative group). Even more dramatic is the fact that of the seven displacement patterns found in the Displacement group, five were *perfect* patterns. No *S* in any of the other groups exhibited a perfect pattern, with the exception of a single case in the Nonreinforcement group. In other words, 50% of *Ss* in the Displacement group met the most strictly defined displacement criterion, in contrast to less than 1% of *Ss* in all other groups combined! This analysis appears to confirm the notion that response displacement occurred as a function of the conflict treatment employed.

[4] The following groups were taken from the Generalization study: *Nonreinforcement* (*Ss* were not subjected to experimental treatment and served as a control group); *Positive* (*Ss* received three trials of positive reinforcement, i.e., successful solution, on the same training puzzle); and *Negative* (*Ss* received three trials of negative reinforcement on the same training puzzle).

SUMMARY

An experiment on preference displacement was designed along the lines of Miller's avoidance-avoidance conflict model, and adapted to a puzzle-playing situation with children. Ten *Ss* were negatively reinforced on their attempts to solve the two puzzles representing the extremes of a five-puzzle series which constituted a shape similarity dimension. *Ss'* differential preferences for the five puzzles were obtained subsequently through paired comparison presentations.

Using a group of 10 different *Ss*, it was shown that a single negative reinforcement trial applied at one extreme of the puzzle similarity dimension generated an avoidance gradient; and it was assumed that a similar gradient would be produced if an identical treatment were applied at the opposite extreme of the similarity dimension. On this basis, it was expected that the most likely resolution of the conflict between the two incompatible avoidance tendencies resulting from the double negative reinforcement applied in the displacement experiment would be at the point where the two gradients intersect; i.e., response peak was expected around the central stimulus of the similarity dimension. This expectation was supported by the results, in that a displacement preference curve was obtained: maximal preference was given to the intermediate puzzle, while the puzzles at the extremes of the dimension were the least preferred. It was also found that this treatment condition produced proportionately more individual displacement patterns than did other treatments not involving the conflict condition.

REFERENCES

BRADLEY, R. A., & TERRY, M. E. Rank analysis of incomplete block designs. I: The method of paired comparisons. *Biometrika*, 1952, **39**, 324–345.

GEWIRTZ, HAVA B. Generalization of children's preferences as a function of reinforcement and task similarity. *J. abnorm. soc. Psychol.*, 1959, **58**, 111–118.

LATCHA, R. Tests of significance in a 2 x 2 contingency table: Extension of Finney's table. *Biometrika*, 1953, **40**, 74–88.

MILLER, N. E. Experimental studies of conflict. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. New York: Ronald, 1944. Pp. 431–465.

MILLER, N. E. Theory and experiment relating psychoanalytic displacement to stimulus-response generalization. *J. abnorm. soc. Psychol.*, 1948, **43**, 155–178.

Received January 23, 1958.

PERSONALITY, PERCEPTUAL DEFENSE, AND STEREOSCOPIC PERCEPTION[1]

JOHN M. DAVIS

School of Medicine, Yale University

WHEN stereograms that present a different picture to each eye are shown through a stereoscope, subjects may report seeing the two pictures in alternation, one superimposed on the other, a mixture of both pictures, or only one of the pictures (in which case the other picture is said to be suppressed). Engel (1955) has suggested that so used, the stereoscope may be a valuable tool for research on relationships between perception and personality and has shown that the meaningful content of the stereoscopic stimuli are important determinants in stereoscopic perception. Cantril (1957) has discussed the stereoscope as an instrument for investigating the role of perception in interpersonal relations. Cross-cultural differences of perception were investigated by Bagby (1957), using American and Mexican subjects, who were presented with stereograms consisting of American scenes paired with Mexican scenes. The scenes from the subject's own culture tended to be perceptually predominant.

This study is an investigation of stereoscopic perception, using an experimental design similar to those employed in tachistoscopic studies of perceptual defense. In the first experiment, some of the variables involved in stereoscopic suppression were investigated, while the second experiment demonstrated differences in the stereoscopic perception of normals and schizophrenics.

EXPERIMENT I

This experiment investigated the effects on stereoscopic perception of the familiarity and emotional content of words, as measured by the Thorndike word count and word association time, respectively. The hypotheses to be tested, which were based on the results of tachistoscopic experiments, were: (*a*) that words with a low frequency of use would be suppressed more than words with a high frequency of use; and (*b*) that words with strong emotional content would be suppressed

[1] I would like to express my appreciation for help in preparation of this paper to E. Engel, A. Alderstein, T. Cornsweet, N. V. Bolin, D. Terry, M. Varley, H. Barry, C. Slack, K. Pribram, A. Lemoine, D. E. Hunt, and A. K. Myers.

more often than words with little emotional content.

Method

Stereograms were shown to 30 male volunteer college students through a Keystone Model 40 prism stereoscope. Subjects (*Ss*) adjusted the slide holder so that a stereogram, consisting of a horizontal line in front of the left eye and a vertical line in front of the right eye, was seen as a cross, giving the proper distance for normal binocular fusion. A card used to cover the stereograms was raised to expose the stimulus word for approximately one second.

Each stereogram consisted of two different typed words, one of which was seen by the left eye and the other by the right eye. Eight stereograms consisted of words of identical word frequencies on the Thorndike (1932) word count, and 16 words were matched with words of different frequencies. Of these 16 pairs, 8 had the more frequent word on the left, while 8 had the more frequent word on the right. The order of presentation of the word pairs was random, as is shown in Table 1. The series of stereograms was presented to one half the *Ss* as it appears on Table 1; the other half of the *Ss* were shown the series with the pairs reversed in position (i.e., words appearing on the left on Table 1 were presented to the right eye, and vice versa). After all *Ss* had seen the 24 stereograms once, the entire series was repeated with the words reversed with respect to the eyes (i.e., all words seen by the right eye in the first presentation were seen by the left in the second showing, and vice versa). The randomization of initial presentation and the systematic reversal serves to control for eye dominance and other variables irrelevant to the present hypotheses.

The *Ss* were asked always to look with both eyes and to report what they saw, spelling words of which they were not sure. If on both presentations of a given stereogram the *S* reported seeing the word on the right, for example, the response to that particular stereogram was said to have been the result of right-eye dominance and was not included in the analysis of the data. After the *Ss* had completed the stereoscopic series, a word association test using the words appearing on the stereograms was administered, and word association time was recorded as a measure of emotional content.

Results

In order to control for the emotional content of stimulus words, the data on pairs of words in which there was at least one word association time of 2.5 seconds or over were analyzed separately from the data on those pairs of words in which the word association time of both words was under 2.5 seconds (see Table 2). When the responses to stereograms containing words of unequal frequency on which both words had association times under 2.5 seconds were analyzed, it was found that 10 *Ss* suppressed[2] more words with the lower frequency of usage according to the Thorndike word count, while only 2 *Ss* suppressed more words with higher frequency of usage. Three *Ss* suppressed equal numbers of words with lower or higher frequencies. By sign test (Siegel, 1957), this result is significant at the .038 level (two-tailed). Thus when word association time is below 2.5 seconds, words of low frequency are suppressed.

On the stereograms on which at least one of the words had an association time of 2.5 seconds or over (i.e., on which the *S* blocked), 5 *Ss* suppressed more words with the lower frequency of usage, 3 *Ss* expressed more of such words, and 4 *Ss* ex-

TABLE 1

THE STIMULUS WORDS AND THEIR WORD FREQUENCIES (The word count of Thorndike (1932) gives the rank order of the words by thousands according to their frequency and range.[a] The lower the word rank (given parenthetically), the more frequent the word in written language. The pairs of words are numbered in the order of their original presentation, the first word of each pair being that which appeared originally on the left.)

1. friend (1) and French (1)	2. clef (18) and blue (1)
3. bottle (2) and breath (2)	4. such (1) and suck (4)
5. hot (1) and cut (1)	6. cockroach (18) and champagne (13)
7. make (1) and fake (1)	8. kiss (1) and hand (1)
9. love (1) and baby (1)	10. gun (1) and pot (1)
11. articulate (8) and chandelier (13)	12. smear (13) and earth (1)
13. clergy (8) and except (1)	14. vulcanize (18) and brilliant (4)
15. bowel (4) and byway (13)	16. arrive (1) and caster (13)
17. enviable (13) and grandeur (8)	18. daughter (1) and mobilize (18)
19. calcium (8) and viscous (18)	20. fad (13) and lay (1)
21. dance (1) and dream (1)	22. cleave (4) and climax (8)
23. dope (13) and ergo (18)	24. defer (8) and bible (4)

[a] When this experiment was being carried out, only the 1932 edition of Thorndike's *Word Book* was available. The word frequencies given here have since been compared to those of the Lorge Magazine Count given in the 1944 edition of the *Word Book*; reanalysis of the data produced no change in the significance levels of the results.

[2] A *suppressed* word is defined here as one which was either inaccurately reported or not reported at all, even though it was exposed before each eye. An *expressed* word refers to a word that was reported correctly in at least one of the viewing conditions. When a word is suppressed, the expressed word, a meaningless mixture of letters, or another word not on the stereogram is reported.

TABLE 2

THE RELATION OF WORD FREQUENCY TO SUPPRESSION

Subject	Stereograms Containing No Words on which *S* Blocked[a]			Stereograms Containing at Least One Word on Which *S* Blocked[a]		
	No. of times more frequent word expressed	No. of times more frequent word suppressed	Sign	No. of times more frequent word expressed	No. of times more frequent word suppressed	Sign
#1	5	0	+	4	2	+
#2	2	0	+	1	1	
#3	1	0	+	1	1	
#4	0	1	−	0	0	
#5	0	1	−	2	1	+
#6	3	2	+	1	0	+
#7	2	1	+	0	1	−
#8	5	1	+			
#9	4	3	+	0	1	−
#10	1	0	+	0	1	−
#11	1	1				
#12	1	1				
#13	1	1		1	1	
#14	4	0	+	1	0	+
#15	1	0	+			
#16				1	0	+
#17				1	1	

Note.—This table shows the total number of responses of all *S*s after eye dominant responses have been deleted. Where no entry appears for an *S*, all of his responses in that category were eye dominant by this definition.

[a] "Blocked" refers to word association times of over 2.5 seconds.

pressed and suppressed equal numbers of the less frequent words (see Table 2) ($p = .73$, two-tailed sign test). The relationship between word frequency and suppression thus does not hold for word pairs in which at least one member has strong emotional content for the *S*, as indicated by long association time.

To see whether a relationship existed between word association time and word frequency, words with association times of 2.5 seconds or over were divided into two groups according to frequency, the first including words of Thorndike rank from 1,000 to 4,000, and the second those ranked from 8,000 to 18,000. It was found that four *S*s blocked more often on words in the category of higher frequency, while six *S*s blocked more often on the less frequent words. A sign test performed on these two groups yields a p of .75 (two-tailed). This experiment does not, therefore, show a relationship between word frequency and word association time by this criterion.

It was found, however, that word association time was related to suppression on a stereogram. Nine *S*s suppressed the majority of words on which they blocked in the word association test, three *S*s suppressed and expressed equal numbers of these

TABLE 3

THE RELATION OF WORD ASSOCIATION TIME TO SUPPRESSION

Subject	No. of Times Words with Association Times of 2.5 Seconds or Over were Expressed	No. of Times Words with Association Times of 2.5 Seconds or Over were Suppressed	Sign
#1	1	5	−
#2	1	1	
#3	1	1	
#4	0	1	−
#5	0	3	−
#6	0	1	−
#7	1	0	+
#9	1	1	
#10	0	1	−
#13	0	2	−
#14	0	1	−
#16	0	1	−
#17	0	2	−

Note.—The data given here constitute all of the responses to stereograms containing words on which the *S* blocked, except those classified as eye dominant responses.

words, and only one *S* expressed the greater number of them (See Table 3). This difference is significant at the .022 level (two-tailed sign test).

It was not possible to randomize the positions of the words with long association times, because they were not determined until after the word pairs had been presented. Therefore a study was made of the patterns of suppression responses. All *S*s clearly showed eye dominance in the majority of their responses. Of the *S*s who suppressed one or more words with emotional content, 6 more often suppressed words which appeared first on the eye dominant side, while 5 more often suppressed words first appearing on the nondominant side.

EXPERIMENT II

Method

Six stereograms were shown to 5 male and 5 female psychotics, patients in a private mental hospital, and to 10 normal *S*s, 6 male and 4 female, medical patients. Both groups of *S*s were in the lower middle class between 20 and 50 years of age; however, the normal group averaged about 10 years younger than the psychotic group.

Each stereogram consisted of two different pictures cut from popular magazines, arranged in such a way that the meaningful parts of each picture stimulated roughly corresponding parts of the two retinae. The series of stereograms was shown to the *S*s twice; in the second presentation the pictures were reversed with respect to the eyes as a control for eye dominance. The stereograms were presented to the *S* for a period of one minute. The *S*s were asked to describe what they

TABLE 4

DESCRIPTIONS OF THE STEREOGRAMS IN EXPERIMENT II

	Picture A	Picture B
1.	A young woman's face	A small figurine
2.	A man's face	A snake
3.	A bowl of Rice Crispies	A Negro man kissing a white girl
4.	A man's head	One man stabbing another with a knife
5.	A man's head	Two women and a man kneeling in prayer
6.	A girl and man in a garden	Two girls and a man in a bathing suit

saw, and their responses were recorded verbatim. Table 4 gives short descriptions of the stereograms.

Responses were classified in five categories. Category 1 (superimposition) consisted of those responses in which the *S* reported seeing both pictures either overlapping or superimposed on each other. Category 2 (alternation) consisted of responses in which the *S* reported seeing the pictures in alternation. Sometimes the pictures would alternate rapidly, and sometimes one picture would slowly fade to be replaced by the other. In Category 3 (suppression) were classified responses in which the *S* saw the same picture on both presentations. In Category 4 (admixture) were placed responses in which he saw an unusual composite of both pictures or something not presented in either picture. The fifth category consisted of eye dominant responses, and was not included in the analysis of differences between psychotics and normals.

Results

On the basis of an unpublished preliminary investigation using psychotics and normals (college students) in a similar testing situation, it was predicted that normals would make more responses in Categories 1 and 3, and that psychotics would give more responses in Categories 2 and 4. When one-tailed Mann-Whitney tests were performed, it was found that psychotics did score significantly higher on Category 4. Normals scored significantly higher on Category 1 and almost significantly higher on Category 3 (see Table 5).

Under the conditions of this experiment, it was clearly evident that different *S*s reported seeing different things even though the stimuli were identical. For example, many *S*s reported seeing a bowl of Rice Crispies and a woman on Stereogram 3, and thus did not report seeing the socially taboo Negro man who was kissing the woman. Three psychotics and no normals reported seeing a baseball game on Stereogram 4. It seems likely that they did actually see the knife fight but interpreted it as a baseball game; that is, they were sensitive to the aggression but denied the aggressive content of the picture.

TABLE 5

COMPARISON OF NORMAL AND SCHIZOPHRENIC SUBJECTS ON FOUR RESPONSE CATEGORIES

Category	Schizophrenics		Normals		p (one-tailed)
	Mean Frequency of S_a' Responses	R_s	Mean Frequency of S_a' Responses	R_n	
1	2.1	78	5.4	132	.025
2	4.0	101.5	2.2	108.5	0
3	1.3	84	3.6	126	.06
4	2.3	131	.6	79	.03

DISCUSSION

Gating mechanisms have been shown to be important in some areas of neural function. For example, Galambos (1956) has shown that auditory nerve discharge is reduced or abolished by electrical stimulation of the floor of the medulla. Hernandez-Peon, Scherrer, and Jouvet (1956) have further observed in cats that a reduction in the auditory response of the dorsal cochlear nucleus occurs with the presentation of a distracting visual stimulus (a mouse), an olfactory stimulus (fish odor), or an electric shock to the forepaw. There is evidence that such neurophysiological gating effects may also be present in vision, olfaction, and cutaneous sensation.

The body of evidence indicating the neurophysiological importance of gating mechanisms would suggest that stereoscopic suppression as well as tachistoscopic perceptual defense may be conceptualized in such terms. The gating mechanism employed in these perceptual operations may be similar to that involved in denial and the other defense mechanisms. In the stereoscopic situation the *S* makes a choice between two stimuli, a choice that is presumably unconscious since he is not aware of the gated-out stimulus. The stereoscopic technique may thus provide an experimental instrument of some importance for the investigation of psychological defense.

SUMMARY

The present paper reports on two experiments on stereoscopic perception. Thirty subjects were shown 24 stereograms, each of which presented a different word to each eye. Association times for the words presented were measured for all subjects as a measure of emotional content, and Thorndike word count ratings were available. Words with low frequency of usage and words with long association time tended to be suppressed. In the

second experiment, some observations made on psychotics and normals indicated that the stereoscopic perception of psychotics differed significantly from that of normal subjects. It is suggested on the basis of this preliminary evidence that the stereoscopic technique might prove experimentally useful, particularly in the study of psychological defense.

REFERENCES

BAGBY, J. W. A cross-cultural study of perceptual predominance in binocular rivalry. *J. abnorm. soc. Psychol.*, 1957, **54**, 331–334.

CANTRIL, H. Perception and interpersonal relations. *Amer. J. Psychiat.*, 1957, **114**, 119–126.

DAVIS, J. M. Four studies in personality. *Education Testing Service*, RM-57-6, 1957.

ENGEL, E. Meaningful content in the study of rivalry and fusion—preliminary observations. Unpublished manuscript, Dept. of Psychology, Princeton University, 1955.

GALAMBOS, R. Suppression of auditory nerve activity by stimulation of efferent fibers to cochlea. *J. Neurophysiol.*, 1956, **19**, 424–437.

HERNANDEZ-PEON, R., SCHERRER, H., & JOUVET, M. Modification of electrical activity in cochlear nucleus during "attention" in unanesthetized cats. *Science*, 1956, **123**, 331–332.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.

THORNDIKE, E. L. *A Teacher's word book of 20,000 words*. New York: Columbia Univer. Press, 1932.

THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words*. New York: Columbia Univer. Press, 1944.

Received February 6, 1958.

THE RELATIONSHIP BETWEEN GSR AND SPEECH DISTURBANCES IN PSYCHOTHERAPY

DAVID M. PANEK AND BARCLAY MARTIN

University of Wisconsin

THE development of procedures for obtaining continuous measures of momentary fluctuations in anxiety level is important for research in psychotherapy. Two such methods have been employed recently. Mahl (1955, 1956a, 1956b) quantified expressive aspects of patient speech, including such categories as "Ahs," "Repetitions," "Sentence Corrections," "Stutters," etc., and found in an analysis of one case that these speech disturbances increased when the patient talked about content which had been independently assessed as anxiety arousing. Dittes (1957a, 1957b) recently employed the GSR as a measure of anxiety and found in a study based on one patient that GSRs accompanying embarrassing sex statements tended to extinguish as psychotherapy progressed. In another study, also based on one patient, he found that GSR activity increased when the therapist became less permissive. If both speech disturbances and the GSR reflect momentary changes in anxiety level, or somewhat more broadly speaking, emotional arousal, one would expect the two measures to covary in a systematic manner, and such is the general hypothesis of the present study.

This research was also concerned with investigating more specific aspects of the hypothesized relationship between the GSR and speech disturbances. For example, it was thought that the category of Repetitions, defined in terms of the unnecessary repetition of words or short phrases, may reflect the immediate disrupting effects of anxiety, whereas Ahs may reflect a more controlled and effective defense against the disrupting effects of anxiety. If such were the case, one would expect Repetitions to be maximal at the point of greatest emotional arousal as indicated by the GSR. Ahs, on the other hand, may indicate the person's attempt to recover from the disrupting effects of anxiety and would accordingly be maximal at a point displaced in time from the point of maximum emotional arousal. This possibility was further suggested by Mahl's (1955) empirical finding that the frequency of occurrence of Ahs tended to be negatively correlated with the occurrence of Repetitions among subjects.

METHOD

Subjects. Four interviews from each of four psychotherapy clients seen by psychologists at the University of Wisconsin provided the basic data for this study. The four clients (three males and one female) were between 19 and 31 years of age. Two had recently begun psychotherapy; two had been in psychotherapy for over a year. Synchronized GSR and tape recordings had been routinely obtained for these clients for most of their therapy sessions. For purposes of this research, four interviews were selected for each client from those for which there were technically satisfactory GSR and tape recordings. The criterion for this selection was the presence of a

relatively large number of scorable GSR dips as defined below. In all parts of the research, analysis of the GSR data was kept independent of the speech disturbance analysis.

Instrumentation. The GSR apparatus was an AC circuit designed by Grant (1946), which used dry silver electrodes attached to the palmar surfaces of two fingers and fed into a G.E. Photoelectric Recorder. At the beginning of therapy, the therapist tapped a telegraph key three times, which simultaneously introduced audible clicks on the tape recorder and made a mark on the GSR record. At the end of the session, the therapist again tapped the telegraph key to provide a final synchronizing signal. There were no apparatus manipulations to concern the therapist between these signals at the beginning and end of the session.

The speech disturbance measures. The most frequently occurring speech disturbances in Mahl's work were Ahs, Repetitions, and Sentence Corrections. In the present study, the speech disturbance scores were obtained directly from the tape recordings rather than from typed transcripts, and preliminary work indicated that Sentence Corrections could not be reliably scored in this way. Accordingly, scores were obtained for only Ahs and Repetitions. Ahs were scored whenever the definite "ah" sound occurred. Repetitions were scored when there was a serial repetition of one or more syllables or one to two words.

The 16 interviews were scored for the number of Ahs and Repetitions occurring in each 30-second interval beginning at the initial synchronizing signal. This was accomplished by connecting a footswitch via a Hunter Timer to the Revere tape recorder so that after a depression of the foot pedal, the tape recorder played for exactly 30 seconds and stopped. The interviews were then played through a second time, and the time that the client and the therapist spent talking in each 30-second interval was obtained by using two time clocks. The speech disturbance scores were divided by the time the client spent talking to yield a rate of speech disturbance as the final score used in statistical analyses.

The GSR measures. The selection of the specific GSR dips to be used in this analysis was made after all interviews had been scored for speech disturbances. The GSR records were marked off in 30-second intervals corresponding to the same intervals used for the speech disturbance scoring. The records were then examined for distinct "dips," that is, those dips relatively isolated in time from preceding or subsequent dips. For purposes of investigating the relationship between GSR and speech disturbance, it was decided to divide the intervals around the GSR dip in the following manner: The 30-second interval in which the dip happened to fall was called the 0 interval, the two 30-second intervals preceding the 0 interval were classified as -1 and -2, and the two intervals following were classified as $+1$ and $+2$.

It was not possible to obtain a sufficient number of distinct and isolated GSR dips, and some compromises were made. The following rather arbitrary criteria were used in selecting usable GSR dips:

1. The dip must have a magnitude of at least 2200 ohms.
2. A 30-second interval was not scorable for speech disturbances if it occurred within two intervals of a succeeding or preceding dip other than the dip being studied.
3. There must be at least three scorable 30-second intervals including the 0 interval associated with a dip. This meant that for some dips, speech disturbance scores were not obtained for one or two of the five possible intervals.
4. A dip was used when other dips were within the surrounding intervals if the selected dip was at least twice as large as those surrounding it and if the dips tended to increase in size approaching the 0 interval, suggesting a high point of emotional arousal.
5. All intervals in which the client spent less than five seconds talking were considered unscorable because the rates of speech disturbances would be based on too small a sample of client talk.

The relationship between the GSR and speech disturbances was further examined by selecting from each interview on the basis of the GSR record the five continuous minutes of maximal GSR activity and the five continuous minutes of minimal GSR activity. One interview was not usable because the GSR activity was consistently high throughout. Thus, the speech disturbance scores used in this comparison were based on the average score for the four interviews for three clients and three interviews for one client.

RESULTS

All speech disturbances were scored by one judge. The reliability of scoring was assessed by having a second judge independently score the first 20 minutes of two interviews. The correlations between the two judges' scores on the 40 30-second intervals involved in each of the two interviews were .87 and .88 for Ahs and .65 and .78 for Repetitions. As an additional reliability check, a third interview was carefully transcribed after the two judges had scored it, with great care being taken to get all Ahs and Repetitions in the typed transcript. The correlations of the two judges' scores with scores obtained from the typed manuscript were .98 and .90 respectively for Ahs and .94 and .76 respectively for Repetitions. It

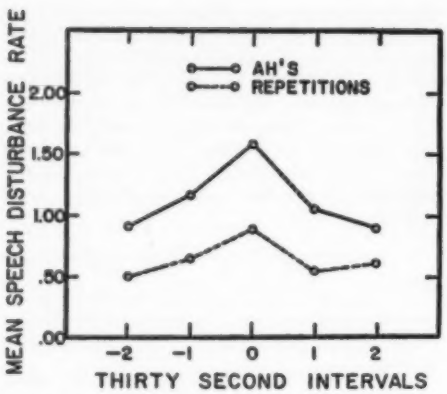


FIG. 1. SPEECH DISTURBANCE RATE AS A FUNCTION OF INTERVALS AROUND GSR DIP

was concluded that the procedure of scoring Ahs and Repetitions directly from the tapes had adequate reliability with the Ahs being somewhat more reliably scored than Repetitions.

Figure 1 shows the mean rate of Ahs and Repetitions as a function of the 30-second intervals surrounding the GSR dip. It is apparent that both Ahs and Repetitions reached a maximum during the 0 or dip interval. An extension of Alexander's test for trend as developed by Grant (1956) was employed to see if the trends manifested in these curves were significantly different from chance. Summary tables for the analyses of variance with respect to Ahs and Repetitions are shown in Tables 1 and 2. The dependent variable used in these analyses was always the average score for a client based on the data for all four interviews, which then yielded four scores representing the four clients at each of the five intervals. For Ahs there is a significant quadratic component, which indicates that the elevation of the curve in the center compared to the more extreme intervals is probably not due to chance. For the Repetitions there are significant quadratic and quartic components, which indicates that it is unlikely that the central elevation as well as the flattening at the extreme intervals is a chance finding. There is no evidence in support of the notion that Ahs and Repetitions would be maximal at different points in time with respect to the GSR dip; both measures reach a maximum at the 0 interval.

One result of scoring the speech disturbances in the intervals around the GSR dip by the present procedure was that the average speech disturbance

TABLE 1

SUMMARY OF TREND TESTS FOR MEAN RATE OF AHS

Source	*df*	Mean square	*F*
Over-all			
Slope	1	.010	0.7
Quadratic	1	.886	16.1*
Cubic	1	.020	0.9
Quartic	1	.400	4.4
Between Individual			
Slope	3	.013	—
Quadratic	3	.055	—
Cubic	3	.023	—
Quartic	3	.090	—

* Significant at .05 level.

TABLE 2

SUMMARY OF TREND TESTS FOR MEAN RATE OF REPETITIONS

Source	*df*	Mean square	*F*
Over-all			
Slope	1	.006	0.7
Quadratic	1	.155	11.1*
Cubic	1	.050	1.2
Quartic	1	.156	12.0*
Between Individual			
Slope	3	.884	—
Quadratic	3	.014	—
Cubic	3	.043	—
Quartic	3	.013	—

* Significant at .05 level.

scores were based on different numbers of GSR dips at the five different intervals. Thus, for the intervals running −2, −1, 0, +1, +2, the total number of dips for which there were scorable intervals at these five points were 39, 69, 84, 74, and 42, respectively. These unequal *N*s might result in greater variability in those intervals with the fewer cases but presumably would not affect the means in any constant way. As a check on this possibility, however, another analysis was performed, using the three middle intervals only, with the necessary GSR dips omitted so as to equalize the number of dips involved. This resulted in the scores for the three intervals being based on 59 dips. The quadratic component was still significant for a combined speech disturbance score based on both Ahs and Repetitions, but it did not reach significance for the separate analyses of Ahs and Repetitions. However, the reduction in the number of dips used undoubtedly increased the variability, and also degrees of freedom were lost in the analysis as a result of eliminating the two extreme intervals.

There were no significant differences between mean speech disturbance scores obtained in the five-minute segments of highest GSR activity as compared with those obtained in five-minute segments of lowest GSR activity, although the differences were in the expected direction ($t = 1.24$ for the combined speech disturbance score). Although the differences were small, in 13 of the 15 interviews the mean rate of combined speech disturbances was higher in the high GSR activity segments than in the low GSR segments.

DISCUSSION

The basic hypothesis of this research that speech disturbances would show an increase at points of emotion arousal as defined by the GSR was confirmed, thus providing some integration of the work of Mahl (1955, 1956a, 1956b) on speech disturbances and Dittes (1957a, 1957b) on GSRs during psychotherapy. The more gross comparison of five-minute segments of high and low GSR activity did not yield significantly different speech disturbance scores. There was also no evidence that Ahs and Repetitions were differently related to points of emotional arousal.

A more general implication of these findings is with respect to the significance of the individual GSR dips. Earlier studies (Bixenstine, 1955; Martin & McGowan, 1955; Mowrer, Light, Luria, & Zeleny, 1953) have frequently focused upon changes in general level of skin conductance or sweating as an indicator of anxiety level or emotional arousal. The fact that in this study discrete GSR dips were found to be related to increases in speech disturbances suggests that in psychotherapy sessions these discrete dips do not, by and large, result from extraneous external and internal stimuli that are irrelevant to the process of therapy. Further analyses of these individual GSR dips in relation to other aspects of the therapeutic process would appear to be indicated. An index based upon both speech disturbance measures and occurrence of GSR dips should prove to be a fairly reliable and valid indicator of the momentary changes in anxiety level in psychotherapy interviews.

SUMMARY

The purpose of this study was to investigate the relationship between points of emotional arousal or anxiety as defined by the GSR and the occurrence of speech disturbances in psychotherapy. Synchronized GSR and verbal recordings of four interviews from each of four psychotherapy clients provided the basic data. It was found that speech disturbances were maximal at times coincident with GSR deflections and decreased in frequency on either side of GSR deflections. This research provides additional validity for both speech disturbances and the GSR as measures of momentary changes in anxiety level in psychotherapy.

REFERENCES

BIXENSTINE, V. E. A case study of the use of palmar sweating as a measure of psychological tension. *J. abnorm. soc. Psychol.*, 1955, **50**, 138–143.

DITTES, J. E. Extinction during psychotherapy of GSR accompanying "embarrassing" statements. *J. abnorm. soc. Psychol.*, 1957, **54**, 187–191. (a)

DITTES, J. E. Galvanic skin response as a measure of patient's reaction to therapist's permissiveness. *J. abnorm. soc. Psychol.*, 1957, **55**, 295–303. (b)

GRANT, D. A. A convenient alternating current circuit for measuring GSR's. *Amer. J. Psychol.*, 1946, **59**, 149–151.

GRANT, D. A. Analysis-of-variance tests in the analysis and comparison of curves. *Psychol. Bull.*, 1956, **53**, 141–154.

MAHL, G. F. The use of "ah" in spontaneous speech. Paper read at Eastern Psychological Association, 1955.

MAHL, G. F. Disturbances and silences in the patient's speech in psychotherapy. *J. abnorm. soc. Psychol.*, 1956, **53**, 1–15. (a)

MAHL, G. F. "Normal" disturbances in spontaneous speech: General quantitative aspects. Paper read at American Psychological Association, Chicago, September, 1956. (b)

MARTIN, B., & MCGOWAN, B. Some evidence on the validity of the Sarason Test Anxiety Scale. *J. consult. Psychol.*, 1955, **19**, 468.

MOWRER, O. H., LIGHT, B. H., LURIA, Z., & ZELENY, M. P. Tension changes during psychotherapy with special reference to resistance. In O. H. Mowrer (Ed.), *Psychotherapy: Theory and research*. New York: Ronald, 1953. Pp. 546–640.

Received February 19, 1958.

SOME FACTORS IN THE SELECTION OF LEADERS BY MEMBERS OF SMALL GROUPS

JOHN P. KIRSCHT, THOMAS M. LODAHL, AND MASON HAIRE
University of California, Berkeley

SEVERAL studies have reported the relationship between leadership and the rate of participation. Bass (1949) found a correlation of .93 between ratings on leadership and the amount of participation time in ten-person groups. However, the size of this correlation may be partially the effect of using the subjects as raters and of the prior acquaintance of the subjects in a class. A study by Borgatta and Bales (1956) reports that high ratings on leadership by group members tends to be associated with high rates of interaction-initiation. Moderate but consistently positive correlations between amount of participation and sociometric scores were found by Peterman (1950).

Although such findings are not unexpected because of the generally task-oriented behavior of laboratory discussion groups, participation, by itself, may not be very predictive of leadership. Slater (1955) reports that leadership as rated by other group members coincided with highest participation in only 11 of 20 discussion groups.

Other researchers have attempted to specify the particular kinds of interaction that differentiate the leaders of small groups. Carter, Haythorn, Shriver, and Lanzetta (1951), using four-person groups, examined the interaction in 53 scoring categories. Two of these categories consistently differentiated leaders in three kinds of tasks: "diagnoses situation—makes interpretation" and "gives information on carrying out action." Other categories that showed differences but not as reliably were: "proposes course of action for others", "initiates action toward problem-solving which is continued," and "integrates group behavior." It was also found that in the reasoning task, leaders ask for information or facts significantly more often than nonleaders. Furthermore, Shaw and Gilchrist (1956) found that leader rank and the number of written communications sent were positively related and that the major source of the difference for leaders was communications about organizing the group and giving factual information.

Thus it seems that leaders are persons who organize the group, solicit and integrate contributions, and propose courses of action. A high rate of participation would not necessarily be associated with these behaviors.

The present study was designed to investigate the relationship between amount of participation, frequency of task- and group-oriented interaction, and the selection of leaders by other group members. The general hypothesis was that amount of participation and organizational-integrative interaction are both associated with leader selection but that each may reflect different aspects of the criterion.

METHOD

Two three-person groups met simultaneously in separate rooms and were given identical instructions by the experimenters. Each group discussed a human relations problem for about 20 minutes, reaching a group consensus in that time. The groups were then told that a second meeting would take place immediately in which one person from each group would act as a representative; the two representatives would discuss a concrete problem related to the previously discussed topic. Each group then selected one of the members to act as representative.

Tape recordings of the meetings were scored for the amount of time each participant talked. Also, each meeting was scored with a system of eight categories based on Bales' (1950) system. In general, the unit of analysis was one simple statement. The three relevant categories and their definitions follow:

D. Gives suggestion (any statement which proposes a course of procedure but is *not* simply an expression of opinion, e.g., "Let's have each person give his solution"—not, "I think teachers should unionize").

E. Asks for suggestion, opinion, fact, (any request which is not for repetition or rhetorical purposes, e.g., "What do you think?" or "Would you agree that. . .").

F. Sums up, integrates (any statement attempting to organize the points covered in the discussion, or to bring together diverging opinions, or to re-word or make explicit opinions of others, e.g., "We seem to agree that strikes are bad"; "These restrictions would qualify what we said earlier").

These three categories will be referred to as "DEF." They define group-oriented, organizational kinds of interaction and are similar to the categories that differentiated leaders in the Carter et al. (1951) study. It was hypothesized that the interaction scored in these categories would distinguish those members who would be chosen by the group as representatives from those not chosen.

Data were obtained on 26 three-person groups.

Two of the groups were discarded because of incomplete interaction records and two were also discarded because they used random devices (flipping coins, etc.) to select the representative.

RESULTS

For the total sample of 66 *S*s, data on time-talked were put into percentage form for each meeting. Those members who were chosen as representatives talked an average of 44.8% of the meeting time; nonrepresentatives talked an average of 27.6% of the time. This difference yields a t of 5.2, significant beyond the .001 level. The average number of DEF interactions per meeting was 12.6 for representatives and 6.1 for nonrepresentatives. However, the variances are heterogeneous in this case, so a median test was used, yielding a chi square of 8.79, significant beyond the .01 level. Thus, both on measures of amount of participation and DEF interaction, those members chosen as representatives were higher.

The time-talked measure correlated with DEF scores +.39, p less than .01. Although a moderate, positive relationship exists between participation and DEF, the two variables do not measure the same thing. When the DEF scores were put into percentage of total DEF per group, the correlation between DEF and time-talked is +.44, not significantly higher than the former correlation.

Since the measure of leadership was a discrete variable, point-biserial correlations were computed to determine the amount of relationship between leader choice and the two interaction measures. It was found that the time-talked measure correlated with leadership +.543; for DEF scores, the point-biserial correlation coefficient was +.527. Both of these correlations are significant beyond the .01 level.

A weighted combination of the two interaction variables was made up using the discriminant function (McNemar, 1955, pp. 210–211). The relative weights were .26 for time-talked and .16 for DEF. This combination yields the best discrimination between the distributions of scores for representatives versus nonrepresentatives. When composite scores were made up using these weights, they correlated with leadership +.632.

If the selection of leader is examined group by group, the predictive power of the interaction measures can be determined and compared. That person who talked the most was chosen as representative in 14 of the 22 groups. Since the chance level of selection is one out of three, this represents a highly significant ($\chi^2 = 9.15, p < .01$) prediction, if not too reliable in a practical sense. In those groups where one person dominated the discussion, i.e., talked over 50% of the time ($N = 8$), time-talked accounts for the representative in seven of the eight cases.

In 14 of the 22 cases, the person with the highest DEF score was chosen as representative. In two of the negative cases, two members were tied in DEF score, and one of the two was chosen. There are 10 cases in which the highest DEF score does not coincide with the highest time-talked. Thus, it is apparent that the two variables do not account for the same thing.

If the dominant person is predicted in those cases in which one person talks over 50% of the time, and where this is not applicable, DEF scores are used, 16 of the 22 cases are accounted for. This method of combining the two variables yields the best practical prediction that can be made from the data.

Although one of the DEF categories may be more crucial than the others, none of the three was found to predict as well by itself as the combined score. It may be that there are particular kinds of groups or situations in which different types of integrative behaviors are particularly appropriate, but these refinements are not possible with the present data.

The other categories used in scoring the interaction yielded no significant differences between representatives and nonrepresentatives. These categories included agreement, positive social interaction, tension release, disagreement, and hostility.

DISCUSSION

The major problem in this experiment is the isolation of variables that can be used to account for the perception of leadership by group members. The data show that amount of participation and DEF interaction are significantly related to leadership choice.

Participation and DEF interaction are neither independent nor highly correlated. They show a moderate, positive relationship. Conceptually, we might think of the group discussion situation in functional terms. The task set by the experimenter defines the goal of the group. Analytically, the process of reaching this goal can be broken down into several components, as is done by Bales (1950). One set of problems involves the production of relevant ideas by the group members; a second set is concerned with the organization and integration of these ideas into a solution that is acceptable to the group. If it can be assumed that the perceptions of the group members of each other with respect to leadership center around the group's functioning, then it appears that the relationship between these two aspects of problem-solving is not a simple one. Within the area of perceived leadership are two subregions: amount of participation and organiza-

tional behavior. In general, these two overlap partially, but it appears from our data that the relative importance and the amount of overlap varies from group to group. In some groups, it appears that any person who takes over the task and produces something will be chosen as leader. In other groups, ideas may be plentiful and the problem is to work out a mutually acceptable result. While this scheme makes difficult a simple prediction of leader choice, it appears to do more justice to the data than a monolithic concept of leadership in small groups.

Another way to state this interpretation is in terms of the differentiation of roles in small groups. Where one member talks a great deal, he is perceived as an appropriate representative, but in a more equal participation situation, group-centered, integrative behavior is relatively more important. In the latter case, persons are differentiated with respect to functioning in synthesizing the group. Even in this case, however, there is undoubtedly a minimum amount of participation necessary before the group-oriented behavior is noticed.

It must be remembered that these groups met only once and that the members were unacquainted with each other prior to the meeting. The representatives are leaders only in the narrow sense that they are empowered to act for the group. While these results may not pertain to a wide range of leadership situations, many short term, problem-solving groups are found that are similar to the experimental situation. Furthermore, the results with respect to DEF interaction seem to corroborate the findings of Carter et al. (1951) regarding the types of interaction which are important in leader behavior. This confirmation includes Carter's negative findings also: the expression of agreement, "personal feelings," and disagreement did not reliably differentiate emergent leaders from nonleaders. In addition, the results on amount of participation tend to confirm the importance of this variable, as has been found by other experimenters.

In summary, the general picture of a small group leader which appears to be common to various researches shows a group member who tends to have a high rate of participation in the discussion; he is task-oriented, attempts to specify the problem, to suggest courses of action, to seek out the members' contributions, to integrate these and to propose solutions in the attempt to secure consensus in the group.

REFERENCES

BALES, R. F. *Interaction process analysis.* Cambridge: Addison-Wesley, 1950.

BASS, B. M. An analysis of the leaderless group discussion. *J. appl. Psychol.*, 1949, **33,** 527–533.

BORGATTA, E. F., & BALES, R. F. Sociometric status patterns and characteristics of interaction. *J. soc. Psychol.*, 1956, **43,** 289–297.

CARTER, L., HAYTHORN, W., SHRIVER, B., & LANZETTA, J. The behavior of leaders and other group members. *J. abnorm. soc. Psychol.*, 1951, **46,** 589–595.

MCNEMAR, Q. *Psychological statistics.* New York: Wiley, 1955.

PETERMAN, J. N. Verbal participation, its relation to decision satisfaction and the leader function in decision making groups. Ann Arbor: Univer. of Michigan Conference Research Project, 1950.

SHAW, M. E., & GILCHRIST, J. C. Intra-group communication and leader choice. *J. soc. Psychol.*, 1956, **43,** 133–138.

SLATER, P. E. Role differentiation in small groups. *Amer. Sociol. Rev.*, 1955, **20,** 300–310.

Received December 27, 1957.

YIELDING, AUTHORITARIANISM, AND AUTHORITARIAN IDEOLOGY REGARDING GROUPS[1]

EUGENE B. NADLER

Case Institute of Technology

THIS investigation was prompted by Asch's (1952) discussion of the personal and social significance of yielding to group pressure. For the person, yielding requires the inhibition of overt and perhaps implicit responses to objective situational requirements. For the group, malignant processes may arise and develop, unchecked by those members in whom goals of personal security predominate over group goals. In the present study, the view was taken that both meanings of yielding are assigned positive value in two separate ideologies concerning the group-individual relation. One of these is termed *conformity,* which includes the value of submission of the individual to the demands of the group, even if so doing requires the inhibition of response to the objective situa-

[1] This article is based in part on a doctoral dissertation submitted to the Department of Psychology, Western Reserve University, February, 1957. The author wishes to thank George A. Albee for his valuable aid and guidance as chairman of the doctoral committee.

tion. The other is *rugged individualism*. Among other features, this ideology countenances the decay of group structures which may result from the valued expression of egocentric motives. Asch's discussion led to the hypothesis that the extent to which individuals yield is related to the degree of their acceptance of either of these two ideologies.

Previous research also suggested the more inclusive hypothesis that yielding, as well as conforming and rugged individualistic ideology, could be viewed within the conceptual framework of the authoritarian personality. Frenkel-Brunswik (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950, Ch. 11) reports that in interviews, people scoring in the highest quartile of the E scale showed acceptance of the idea of survival of the fittest, hero worship of acquaintances, a view of the world as a jungle, and exploitive manipulative opportunism, i.e., rugged individualism. The *same* people showed conventionalism and the desire for status-acceptable qualities in their friends, i.e., conformity. However, her account does not reveal the extent of relationship between the two patterns, nor how they influence behavior in a concrete situation.

METHOD

Group pressure was simulated by announcing bogus group judgments to small groups of *Ss* whose apparent task was to match one of three lines of variable length with a standard line (Asch, 1952, Ch. 14). This method was proposed by T. Rice,[2] and is similar to that independently announced by Crutchfield (1955) shortly after the present study was begun. A more complete description of the method, and the precautions taken to maintain the deception, may be found in Nadler (1957).

Questionnaires were used to measure the other variables, which were: the ideology of conformity (IC), the ideology of rugged individualism (RI), ethnocentrism (E), and authoritarianism (F). The California E and F scales (Adorno et al., 1950) were used to measure the latter two variables. Similar Likert type scales were used to measure conformity and rugged individualism. Reliability and internal consistency of IC and RI had been determined beforehand in 107 college students.

Items in the IC scale stressed the value of compliance with the demands of common reference groups and the necessity for submerging the self in the group. Punishment for dissent was both implied and explicated, as was the hopelessness of dissent. Items in the RI scale denied the importance of social forces, asserting both the undesirability and the impossibility of groups having any decisive influence upon individuals. Other items raised narrow self-interest to the level of a virtue and denied social motives.

Since the original F scale contained many items with conforming or rugged individualist content, it was necessary to omit such items from the version used here, particularly since some of the items had been used in the IC and RI scales.[3]

The questionnaires were administered to 91 college students, both day school and evening. About one week later, 70 of these participated in the experiment.

TABLE 1

STATISTICAL PROPERTIES OF THE MEASURES

Property	Measure				
	IC	RI	E	F	Yielding[a]
Reliability	.76	.62	.85	.74	.52
Mean	3.60	4.21	3.28	3.70	2.84
SD	.86	.79	.99	.88	2.52
N	91	91	71	91	70
No. of items	20	20	14	14	12

[a] Mean and *SD* of this measure refer to the full scale, each item scored pass-fail; means and *SD*s of the other measures refer to each item, scored 7-1 on an agree-disagree continuum.

TABLE 2

INTERCORRELATIONS AMONG THE MEASURES

Measure	RI	E	F	Yielding
IC	.40	.55	.68	.30*
RI		.56	.63	.29*
E			.61	.30*
F				.48

* Significant beyond the .025 level. All others significant beyond the .01 level.

RESULTS

Statistical properties of the several measures are given in Table 1. The correlations between the frequency of yielding and each of the four scales are presented as part of Table 2. Negro, Jewish, and Asian *Ss* were not included in correlations involving the E scale. IC, RI, and E showed slight but significant ($P < .025$, one-tailed test) correlations with frequency of yielding. The correlation with the F scale was .48 ($P < .01$), somewhat higher than Crutchfield's (1955) finding which it

[2] Personal communication from Western Reserve University graduate student.

[3] This version of the F scale included the following items: Item 9, Form 78 (Adorno et al., 1950, p. 226); Item 44, Form 60 (Adorno et al., 1950, p. 248); Items 4, 6, 8, 9, 13, 18, 19, 25, 31, 35, 39, and 43, Forms 45 and 40 (Adorno et al., 1950, pp. 255–257).

replicates. The correlations of yielding with IC and RI were significantly lower than that with F ($P < .05$ in both cases). The correlation with the E scale was not significantly different from the correlation with F, the removal of cases required by this comparison causing the correlation with F to drop.[4]

The intercorrelations among the scales are also presented in Table 2. These show moderate to substantial positive relationships ($P < .01$) among the various measures. As a check against spurious correlation between IC and RI due to similarity of item content, two psychologists sorted the otherwise unidentified items into the proper ideological categories. Both judges correctly divided the items without an error.

DISCUSSION

It will be recalled that the F scale used in this study deliberately omitted items with conforming or rugged individualistic content. The remaining items express largely superstition, projectivity, sex preoccupation, moralism, and destructiveness. Summed up, these imply a weakening of the adaptive functions and lack of confidence in them. The higher correlation of yielding with the F scale than with IC and RI point to the conclusion that central personality variables as revealed by the F scale were more important determinants of yielding than more peripheral ideological variables as revealed by IC and RI. The evidence was equivocal for the E scale, since the drop in correlation occasioned by removal of cases may have been due to the resulting homogeneity of variance. The authors of the F scale seem to have succeeded in their purpose of building an instrument that would measure those underlying personality tendencies responsible for antidemocratic ideology.

The more peripheral nature of the ideologies of conformity and rugged individualism probably accounts for the character of the relationship between them, low positive. Both are authoritarian ideologies, each having anti-individual as well as antisocial qualities. Yet, if they were *simply* dual and interchangeable expressions of authoritarianism, the correlation between them would be as high as their separate correlations with authoritarianism. Since this is not the case, it appears that they function only partly as expressions of authoritarianism, but partly as psychological alternatives. It might be rewarding to study the way these alternatives are used by yielders.

[4] Comparison of r_{12} and r_{13} requires that $N = N_1 = N_2$. In the present instance, certain cases were dropped, and some of the correlations in Table 2 were recalculated. (See McNemar, 1949, p. 124.)

REFERENCES

ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.

ASCH, S. E. *Social psychology*. New York: Prentice-Hall, 1952.

CRUTCHFIELD, R. S. Conformity and character. *Amer. Psychologist*, 1955, **10**, 191–198.

MCNEMAR, Q. *Psychological statistics*. New York: Wiley, 1949.

NADLER, E. B. The ideological correlates of conformity. Unpublished doctoral dissertation, Western Reserve Univer., 1957.

Received January 20, 1958.

OPUS #1

An event of importance for
NEW AUTHORS ONLY

OPUS #1 is the title of a new Philosophical Library 1959 book series which will publish the deserving manuscripts of new authors only.

OPUS #1 has two basic requirements: The author must have something to say and know how to say it.

OPUS #1 will be dedicated to the humanities and will include works in the fields of Art, Literature, Philosophy, Religion, Sociology, History, Psychology and Psychiatry.

OPUS #1 books will be printed on special, fine antique paper, and artistically cloth-bound for library use.

Only complete book-length manuscripts accepted.

Inquire for full details
Philosophical Library, Publishers
Dept. OP-105
15 East 40 St., N.Y. 16, N.Y.

Schizophrenia Manfred Sakel	5.00
Epilepsy Manfred Sakel	5.00
The Analysis of Dreams Medard Boss	6.00
Group Psychoanalysis B. Bohdan Wassell	3.75
The Neuroses and their Treatment Edward Podolsky	10.00
Psychotherapy and Society Wladimir Eliasberg	6.00
The New Chemotherapy in Mental Illness H. L. Gordon	12.00
Principles of Self-Damage Edmund Bergler	6.00
The New Psychiatry Nathan Masor	3.75
Experimental Psychology Ivan Pavlov	7.50

Philosophical Library, Publishers
15 East 40 Street, New York 16

THE AMERICAN JOURNAL OF PSYCHOLOGY

Founded in 1887, The American Journal of Psychology is largely devoted to the interests of experimental psychology, although its pages are open to workers in all fields of scientific psychology. The journal is published quarterly in March, June, September, and December. Subscriptions: $7.00 a year (single numbers $1.85); foreign, $7.50 a year (single numbers $2.00). Subscriptions, back numbers and volumes may be obtained from the Business Office.

Portraits of Psychologists

Offprints of Portraits of Psychologists that have been published in The American Journal of Psychology are now available. Listed in order of their appearance, they are as follows:

1. G. Stanley Hall, Vol. 35, 1924, facing p. 313
2. Edmund C. Sanford, Vol. 36, 1925, facing p. 157
3. James Ward, Vol. 36, 1925, facing p. 449
4. Franz Brentano, Vol. 36, 1925, facing p. 469
5. Edward B. Titchener, Vol. 38, 1927, facing p. 489
6. Margaret Floy Washburn, Vol. 39, 1927, frontispiece
7. Herman Ebbinghaus, Vol. 42, 1930, facing p. 505
8. Wilhelm Wundt, Vol. 44, 1932, facing p. 615
9. Howard C. Warren, Vol. 46, 1934, facing p. 189
10. G. E. Müller, Vol. 47, 1935, facing p. 187
11. Carl Stumpf, Vol. 49, 1937, facing p. 173
12. Kurt Koffka, Vol. 55, 1942, facing p. 157
13. William James, Vol. 55, 1942, facing p. 309
14. Raymond Dodge, Vol. 55, 1942, facing p. 467
15. John A. McGeoch, Vol. 56, 1943, facing p. 1
16. James McKeen Cattell, Vol. 57, 1944, facing p. 133
17. Max Wertheimer, Vol. 57, 1944, facing p. 291
18. Joseph Jastrow, Vol. 57, 1944, facing p. 451
19. Charles S. Myers, Vol. 60, 1947, facing p. 159
20. Pierre Janet, Vol. 60, 1947, facing p. 479
21. James R. Angell, Vol. 62, 1948, facing p. 315
22. Knight Dunlap, Vol. 63, 1950, facing p. 1
23. Edward L. Thorndike, Vol. 63, 1950, facing p. 161
24. Carl E. Seashore, Vol. 63, 1950, facing p. 317
25. Oswald Külpe, Vol. 64, 1951, facing p. 1
26. Clark L. Hull, Vol. 65, 1952, facing p. 497
27. David Katz, Vol. 66, 1953, facing p. 519
28. Walter S. Hunter, Vol. 67, 1954, facing p. 571
29. Harvey A. Carr, Vol. 68, 1955, facing p. 1
30. Louis L. Thurstone, Vol. 69, 1956, facing p. 1
31. Madison Bentley, Vol. 69, 1956, facing p. 169
32. Robert M. Yerkes, Vol. 69, 1956, facing p. 345
33. Lewis M. Terman, Vol. 70, 1957, facing p. 333
34. Karl M. Dallenbach, Vol. 71, 1958, facing p. 1
35. Karl S. Lashley, Vol. 72, facing p. 1
36. John B. Watson, Vol. 72, facing p. 163

Separates—35¢ each

(Please order by number)

The series—$9.00

THE AMERICAN JOURNAL OF PSYCHOLOGY

University of Texas Mezes Hall 102

Austin 12, Texas

New LOGOS Publications in the Behavioral Sciences

Schizophrenia: A Review of the Syndrome. Edited by *Leopold Bellak, M.D.* Over 4,000 studies of the past decade reviewed and integrated. "It would be hard to overestimate [its] importance . . . for practitioners, researchers, librarians, and the many others who need to know . . ."—*Digest of Neurology and Psychiatry*, Jan. '59.
1010 pp., $14.75.

The Rorschach and the Epileptic Personality. By *Delay, Pichot, Lempérière* and *Perse*; introduction by *Arthur L. Benton*, commentary by *Zygmunt A. Piotrowski*. He writes: "In all the vast [Rorschach] literature there is nothing like this book. . . The catholicity of the approach is responsible for much of the value and precision of the conclusions. . . The professional who is interested in applying advances in perceptanalytic diagnosis to neuropsychiatry will find many issues clarified. . . A model of what the Rorschach approach to a diagnostic problem should be."
265 pp., $6.

Myokinetic Psychodiagnosis (M.K.P.). By *Emilio Mira y Lopez*. The M.K.P. Test, widely used in Europe and Latin America, introduced to the U. S. *Gordon W. Allport* says in his foreword: "This is an important book . . . a diagnostic method of large promise . . . fixes our attention on the most neglected area of personality theory . . . underlying dispositions of temperament . . . manifested in expressive movement." M.K.P. permits quantitative analysis as well as psychodynamic interpretation.
186 pp., 73 illus., $6.75.

Booklets for M.K.P. Test Administration, package of 25, $5.

A LOGOS standard text, widely used

Psychology of Personality: Six Modern Approaches. Edited by *J. L. McCary*. "An intellectual feast . . . on the frontier of contemporary thought," writes *Henry A. Murray* of these varied views of personality from *Bellak, Cattell, Klein, McClelland, Mead* and *Sanford*.
383 pp., $6.75.

From:

LOGOS PRESS, • Box 273, Cooper Station, New York 3.

MICROFORM

All journals published by the American Psychological Association are being made available on MICROFILM or MICROCARD.

Psychological Review
American Psychologist
Psychological Bulletin
Psychological Abstracts
Contemporary Psychology
Psychological Monographs
Journal of Applied Psychology
Journal of Consulting Psychology
Journal of Educational Psychology
Journal of Experimental Psychology
Journal of Abnormal and Social Psychology
Journal of Comparative and Physiological Psychology

Available only in volume units; no single issues.

For MICROFILM, order from:
UNIVERSITY MICROFILMS, INC.
313 North First Street
Ann Arbor
Michigan

For MICROCARD, order from:
J. S. CANNER & COMPANY, INC.
Microcard Division
618 Parker Street
Roxbury 20, Massachusetts